

WHO's in charge?

Global health remains firmly on the G8 agenda — for better or worse.

Last weekend's G8 summit in St Petersburg, Russia, may have been notable primarily for its state of paralysis over the latest crisis in the Middle East. However, behind the scenes, the meeting marked further progress in the group's assumption of responsibility for tackling global health problems. This attention is welcome — but the nature of the G8 forum and its exclusion of notable players, such as China, India and Brazil, carries risks.

Infectious diseases were one of the G8's top three declared priorities, alongside energy security and education, and the summit featured the first preparatory meeting of health ministers in the forum's thirty-year history.

A proposal to lure companies back into vaccine research by guaranteeing a market faltered, with Japan and Germany reportedly balking at the cost. But the meeting reiterated its support for financing the Global Fund To Fight AIDS, Tuberculosis and Malaria, and a host of other G8 health commitments, such as eradicating polio and making AIDS antiretrovirals available to all by 2010. There were no announcements of substantial new funding or initiatives, however.

Prompted by the world's failure to stop the spread of H5N1 avian flu, the meeting attempted to address the neglect of disease surveillance and emerging diseases (see *Nature* 440, 6–7; 2006). The spread of avian flu has exposed both the vulnerability of rich countries to diseases that emerge outside their own borders, and their lack of the vaccines or infrastructure needed to resist a human pandemic. In response, a G8 communiqué calls for better international cooperation on surveillance and monitoring, building infrastructure in affected countries, “full transparency by nations in sharing, on a timely basis, virus samples and other relevant information about the outbreaks of diseases” and “encouraging development of next generation influenza vaccines”.

But this is just a wish-list, with no roadmap or funds — although the G8 reminded the world's countries that they should pay up the \$1.9 billion they pledged to the United Nations' avian-flu effort last January in Beijing, something most of them have yet to do.

Those who work in global health will, nonetheless, welcome the G8's explicit recognition of the importance of disease surveillance,

and will hope that political attention at this level will ultimately pay off in the field.

Health issues have gradually grown in prominence at such summits, after AIDS and severe acute respiratory syndrome (SARS) reminded rich countries that they are not immune from emerging diseases. But tackling them requires far more resources than the G8 has so far managed to muster. And the attention of this elite group has had a side effect: the subtle shifting of money and influence away from the World Health Organization (WHO) and other UN programmes towards new organizations and government initiatives.

That shift may be no bad thing. The WHO is bureaucratic and sycophantic to its sovereign member states, and has a track record of failure in its programmes to tackle malaria and AIDS, although it fared better in handling smallpox and the 2002–2003 SARS outbreak. Like the UN Food and Agriculture Organization, which has taken a prominent role in the global response to avian flu, it is underfunded and chronically understaffed in key departments.

But the shift away from the WHO and other multilateral UN agencies carries risks. The G8 does not represent the world's people, and its members' national interests, including commercial ones, are not always aligned with global public health. A G14 or G20, including the leading developing countries, might do better in that regard.

Alongside the Global Fund and other G8 initiatives, the world's efforts to combat disease are also being transformed by private-sector initiatives, most notably those of the Bill and Melinda Gates Foundation, with its focus on late-stage product development.

But these efforts seem small and fragmented next to the challenges they are supposed to address. Future summits need to confront two major questions: how funding can be increased to levels that can do the job, and how, 58 years after the WHO's creation, the international community intends to agree and fund a coherent set of research and control priorities for global health. ■

“The G8 does not represent the world's people, and its members' national interests are not always aligned with public health.”

Safety catch

Risk assessment is a useful environmental tool, but not if it is used as a cover for a deregulatory agenda.

A bulletin issued earlier this year by the White House Office of Management and Budget contains a number of recommendations on how the different parts of the US federal government should go about assessing risk (see page 242). The document, when it is finalized, will have an important bearing on

how regulatory decisions such as environmental rules are made.

The topic of risk assessment sounds arcane but is of vital importance, especially to the United States' poorest communities. The poor have no say in setting the rules but bear the brunt of most environmental threats, including dirty water, polluted air and chemicals left behind on industrial sites. They will suffer the consequences if the balance of risk assessment is shifted in favour of the polluter. And if the current draft is implemented, that's exactly what will happen.

The United States has pioneered the use of quantitative risk assessments, which are now widely used around the world. The National Academies has played a central role in setting the agenda for how



such assessments should be conducted and their outcomes incorporated into the related sphere of 'risk management', whereby regulators and other agencies take action in response to an identified risk.

The proposed bulletin would increase the range of circumstances in which formal risk assessment would be required before government agencies could take action or set regulations. It would also put in place firm guidelines on how these assessments are conducted.

This effort echoes the legislation on risk assessment and cost-benefit analysis that the Republican-led Congress attempted to pass in the late 1990s. That legislation failed, opposed by moderate Republicans such as Sherwood Boehlert, now chair of the House science committee, who rightly saw it as an attempt to stifle the Environmental Protection Agency (EPA), the Food and Drug Administration and other regulators.

That legislation was, at least, a relatively transparent attempt to roll back regulation, which at the time was an important element in the Republicans' political agenda. The call for government to get off the backs of companies and individuals had considerable resonance then, and indeed it still has. But it is an argument that has lost some of its political appeal, and it is certainly not being made in public to support the White House's proposed risk-assessment bulletin.

Some risk assessments done by government agencies do fall short

of reasonable standards. Only last week, a National Academies panel criticized an assessment by the EPA into the chemical dioxin. But these shortfalls could be addressed without tying up the whole government in a set of rules to be administered from the centre by a small, heavily politicized office with few technical staff — the OMB's Office of Information and Regulatory Affairs.

It is not the first time the OMB has sought to reform the regulatory environment through this office, whose recently departed director, John Graham, was a former head of the Harvard Center for Risk Analysis. When it proposed a strict definition of how the science behind regulatory decisions should be peer-reviewed, the National Academies cried foul, and the definition was relaxed.

This time, the OMB has asked the National Academies to review the proposed bulletin, confident that such a review will endorse its technical content. But the bulletin's technical content is not being disputed: what is at issue is its scope, suitability for purpose, cost and the effort that might be wasted in enforcing compliance.

The motivation of Graham, his mooted successor Susan Dudley of George Mason University in Virginia, and indeed of President Bush himself, is not really in doubt. What they want is not better regulation, but less regulation. They should admit as much, instead of hiding their agenda behind the mantra of 'sound science'. ■

Save the lungfish

An Australian dam project threatens a living fossil.

The Australian lungfish, *Neoceratodus forsteri*, has become the focus of an important conservation battle in Queensland (see page 232). State government plans to build a dam on the Mary river could threaten the species' survival, conservationists say.

Far be it from *Nature* to put one species above another, but in this case the authorities making the dam decision need to be aware that there are particular reasons why the lungfish is worth conserving.

The fish, also known as the Queensland lungfish, is native to the Burnett and Mary rivers in southeastern Queensland. It is also found in small numbers elsewhere in the state, but these are largely introductions, and not very successful ones at that. The Burnett is already being dammed, and a second dam, on the Mary River, would effectively destroy the last pristine habitat for the species.

The Australian lungfish is the most scientifically interesting of six known, surviving species of lungfish — a kind of fish that was prevalent in both sea and freshwater during the Devonian period (some 400 million years ago). Like all primitive fishes it has a lung, as well as gills. The immediate (but not very close) relatives of the lungfish include the coelacanth, as well as the ancestors of all land-living vertebrates. This makes the lungfishes 'living fossils' of great value in studying the biology of the earliest ancestors of land animals (see E. B. Daeschler *et al.* *Nature* **440**, 757–763; 2006). Studying the species may provide a unique insight into how our own vertebrate ancestors made the journey from water to land.

But while the other lungfish species, found in Africa and South America, have come to resemble amphibians, the Australian one is more fish-like. Fossils of the species date back 100 million years, to

a time when dinosaurs stalked the land. Its prominent, fleshy fins and heavy scales make it look exactly like a Devonian lungfish from hundreds of millions of years earlier — quite unlike the smooth-skinned, bootlace-finned African lungfishes, for example.

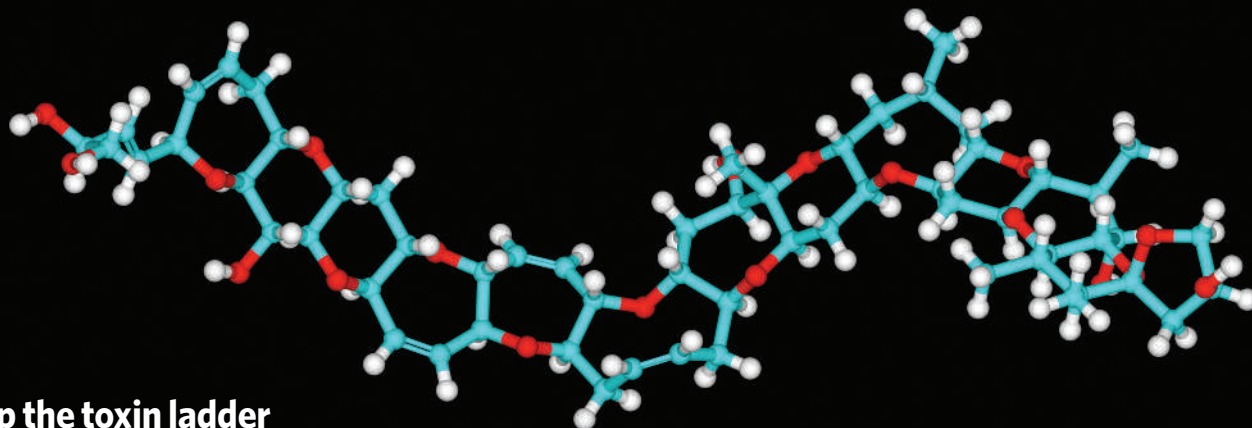
Although the Australian lungfish can swim in deep water as an adult, it tends to forage in the shallows and has an absolute requirement for shallow, slow-flowing, weed-choked water in which to spawn. The permanent flooding caused by the dams will destroy this habitat. There are plans to create a channel to allow fishes to migrate past the Mary River dam, but this does not address the fish's peculiar reproductive requirements. In addition, the introduction of other fishes, such as tilapia, is thought to be threatening the lungfish through predation on its eggs and young.

Scientists and others concerned about the fate of the lungfish will be able to put their views forward during extensive environmental assessments, which both the state and the federal government will now undertake to consider the impact of the current dam proposal — not just on the fish, but on the local ecosystem and economy as a whole. Obviously the impact of the project has to be weighed against the water needs of southeastern Queensland, Australia's fastest-growing region. And evolutionary biologists opposed to the dam must also consider creative ways of ensuring that the fish can survive, in the event that the dam is built.

But the status of the Australian lungfish as rare, distinctive and of profound interest to evolutionary biologists makes its conservation much more important than that of an attractive species of bird, say, which has many living relatives. This distinction should be brought to the forefront in making a decision on the proposed dam. ■

"Study of the species may provide a unique insight into how our own vertebrate ancestors made the journey from water to land."

RESEARCH HIGHLIGHTS



Up the toxin ladder

J. Am. Chem. Soc. doi:10.1021/ja063041p (2006)

Chemists have mastered the making of ciguaterins, the exceptionally large and highly toxic molecules responsible for some forms of seafood poisoning. By increasing the amount of the toxins

available to test, this work could help biologists to probe how the toxins work *in vivo* and find drugs to combat ciguatera poisoning.

Ciguaterins are produced by the dinoflagellate *Gambierdiscus toxicus*, which is eaten by fish

and so in turn is consumed by the occasional hapless human. The toxins present a tantalizing challenge to synthetic chemists both because of the molecules' size and their complex 'ladder like' architecture (pictured).

Masayuki Inoue, Masahiro Hirama and their co-workers at Tohoku University in Japan report the total synthesis of the two most toxic members of the family: ciguaterin and 51-hydroxyCTX3C.

ANIMAL BEHAVIOUR

Love songs

Curr. Biol. **16**, 1311–1316 (2006)

Flying mosquitoes may change the whining tone created by their wings to match those of potential mates, say Gabriella Gibson of the University of Greenwich and Ian Russell of the University of Sussex, UK.

The researchers recorded the wingbeat frequencies of pairs of tethered tropical mosquitoes, *Toxorhynchites brevialpalis*. They found that both mosquitoes in an opposite-sex pair alter their wingbeat patterns until their flight tones are the same. By contrast, the noises made by any same-sex pair diverge in frequency. Males, who have more sensitive antennae, make wingbeat adjustments more quickly than female mosquitoes.

BIOPHYSICS

Enzyme fieldwork

Science **313**, 200–204 (2006)

Steven Boxer and his colleagues at Stanford University, California, have devised a way to fit a voltmeter inside an enzyme, making it possible to probe the electric fields that influence the enzyme's structure and catalytic activity.

Boxer's group showed that changes to the electric field inside an enzyme, induced by mutations in the protein, shift the vibrational frequency of a carbon–nitrogen triple bond in a molecule bound in the enzyme's active site.

The researchers measured the frequency

shift for an inhibitor of the human enzyme aldose reductase, but say the technique should be applicable to many other enzymes. In this case, the observed changes in field strength matched the predictions of computer simulations that included movements of the enzyme's side chains.

CELL BIOLOGY

Death bottled up

Cell **126**, 163–175; 177–189 (2006)

Mitochondria (pictured below) appear to use their membrane folds as bottles for a protein — cytochrome *c* — that can boost programmed cell death. Sequestering the protein protects the organelles' host cells from an untimely end.

That mitochondria release cytochrome *c* to amplify apoptosis was already known; the

papers from Bart De Strooper of the Flanders Interuniversity Institute for Biotechnology in Leuven, Belgium, and Luca Scorrano of the Dulbecco-Telethon Institute at the Venetian Institute of Molecular Medicine, Padova, Italy, and their colleagues explored how this happens.

They show that a soluble form of the protein OPA1 seals cytochrome *c* inside pocket-like folds of the inner mitochondrial membrane. The plug seems to be maintained by the enzyme PARL, which cleaves a membrane-bound form of OPA1 to produce the soluble form. The authors suggest that the actions of other cell-death activators might disrupt the OPA1 stopper, releasing cytochrome *c*.

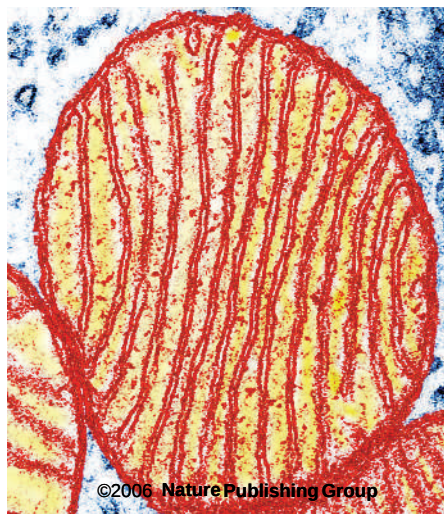
GENETICS

Buffering disease

Nature Genet. doi:10.1038/ng1844 (2006)

A person's susceptibility to genetic disease may be influenced by a class of genes that buffer against the effects of mutations, suggest researchers in Europe.

Andrew Fraser of the Wellcome Trust Sanger Institute in Cambridge, UK, and his colleagues used RNA interference to investigate the interactions between about 65,000 pairs of genes in the worm *Caenorhabditis elegans*. They produced a map of these interactions and found six 'highly connected' genes that together interacted with a quarter of all the other genes tested. If one of these six genes was disabled, the effects of mutations in many



©2006 Nature Publishing Group

ISM/SPL

other genes became worse.

All six genes regulate gene activity by affecting the structure of chromatin, the combination of DNA and proteins that makes up chromosomes

ASTRONOMY

View from the APEX

Astron. Astrophys. **454**, L13–L118 (2006)

Perched at an altitude of 5,107 metres in the Chilean Andes, the Atacama Pathfinder Experiment, APEX, is a 12-metre telescope that began scanning the southern skies in 2005. Its first observations and a description of the instrument have now been published in a special package of 26 letters.

The APEX (pictured right) gathers far infrared and longer wavelength light. So far, its results include the first ever detection of the fluoromethyldinium ion (CF^+) in space, consistent with theories of interstellar chemistry. The telescope is paving the way for the Atacama Large Millimeter Array (ALMA), which will consist of about 50 telescopes of the same size, under construction in the same area.

AIDS RESEARCH

Cutting HIV

PLoS Med. **3**, e262 (2006)

Male circumcision could slash the number of fresh HIV infections in sub-Saharan Africa by some 2 million, and reduce deaths by 300,000, over the next decade.

The estimate follows the first randomized, controlled trial of male circumcision. This showed that the practice cuts HIV transmission from women to men by around 60% (B. Auvert *et al.* *PLoS Med.* **2**, e298; 2005).

Now Brian Williams of the World Health Organization in Geneva, Switzerland, and his

colleagues have used mathematical models to predict the health impact of extending circumcision to all men in the region.

In the second decade, this simple practice could prevent 3.7 million additional infections and 2.7 million deaths, the researchers estimate.

COMPUTING

Dot gain

Appl. Phys. Lett. **89**, 013503 (2006)

The conventional circuitry of a computer chip might one day be replaced by tiny components known as quantum cellular automata (QCA).

Mladen Mitic and his colleagues at the University of New South Wales in Australia have brought this idea a step closer to reality by building a QCA cell from four nanoscale blobs, or quantum dots, of silicon.

A QCA cell stores information by switching between two states — quantum dots do this by swapping electrons with each other. It is expected that such QCA cells, suitably configured, could run a computer algorithm because their states will switch through interactions with their neighbours.

Previous QCA cells have used dots made

from metals or semiconductors. By using silicon, this team have made a cell that is compatible with existing microchip technology.

EVOLUTION

Rise of the radish upstarts

Evolution **60**, 1187–1197 (2006)

Everyone wants their offspring to do well in life, but not at their own expense. Evolutionary biologists studying California's wild radishes (*Raphanus*) have found that one hybrid of two radish species has been so successful in some regions of the state, that it has driven both of its parents to extinction.

Such swamping of existing species by new hybrids is difficult to observe in the field, explain Subray Hegde of the University of California, Riverside, and his colleagues, because it can happen within just a few generations.

The researchers used greenhouse breeding experiments and genetic comparisons to confirm that California's wild radishes, now an important weed species, are indeed the product of a fruitful union between two previous species, *Raphanus raphanistrum* and *R. sativus*.



APEX

JOURNAL CLUB

Albrecht W. Hofmann
Max Planck Institute for
Chemistry, Mainz, Germany

A geochemist suspects that Earth's oldest rocks may have been hidden away.

One of the greatest puzzles in Earth science, and one I find particularly fascinating, is rooted in the deepest part of the mantle. Just outside the liquid core there exists an irregular layer, some 200 kilometres thick, that has baffling

seismic properties.

Seismologists and mineral physicists have pinned their hopes of explaining this D-double prime (D'') layer to a high-pressure mineral phase first reported in 2004, known as post-perovskite. But some geochemists, including this one, favour another model.

Last year, researchers made the surprise discovery that terrestrial rocks are richer in a certain isotope of neodymium, ^{142}Nd , than are meteorites, the bricks from which Earth was built (M. Boyet & R. Carlson *Science* **309**, 576–581;

2005). One way to reconcile the results is if Earth contains a hidden reservoir deficient in ^{142}Nd that balances the measured excess — and maybe D'' is it.

The isotope ^{142}Nd is produced by decay of the relatively short-lived ^{146}Sm , a now-extinct isotope of samarium. Crust concentrates Sm slightly less efficiently than it does Nd, so a primordial crust that formed early in Earth's history, when there was still plenty of ^{146}Sm around, would have a lower Sm/Nd ratio than the mantle and ultimately, therefore, a lower

^{142}Nd abundance.

Boyet and Carlson suggest this primordial crust was subducted and stored at the base of the mantle, providing the reservoir deficient in ^{142}Nd and giving rise to D'' .

I like this model because it lets geochemists have a mantle with separate layers, which helps to explain some other peculiarities of Earth's composition. Previously, this idea seemed to fly in the face of seismic images showing Earth's mantle to be thoroughly mixed, except for D'' , that is.

NEWS

Moving towards a graphene world

Welcome to graphene: the flat carbon sheet with revolutionary aspirations. This thinnest possible pencil-lead shaving has already interested theoretical physicists with its electronic properties, and is predicted to edge aside silicon in the microchips of the future. Now it's ready for its first practical application.

Carbon exists in many forms: 'buckyballs', diamond, nanotubes and graphite to name a few. In 2004, André Geim of the University of Manchester, UK, together with Russian colleagues, created another form — carbon layers just one atom thick, called graphene (K. S. Novoselov *et al. Science* **306**, 666–669; 2004). Physicists have since documented some remarkable properties of this material. But industrial applications have seemed some way off, as graphene has proved difficult to produce on a large scale. It was made by unwieldy methods such as rubbing flecks off a piece of graphite, as Geim did, or boiling up silicon carbide in a vacuum.

Now, Rodney Ruoff and his team at Northwestern University, Illinois, have come up with a way to make large amounts of graphene embedded in a polymer matrix. The researchers start with graphite oxide — graphite with oxygen-containing chemical groups attached. After further chemical modification and treatment with ultrasound, the material separates into layers, and disperses through a solvent in which a polymer such as polystyrene is also dissolved. Chemical reduction removes most of the oxygen groups, and removing the solvent leaves behind graphene sheets crumpled within the solid polymer (see pages 282 and 254).

The resulting material is strong, and electrically and thermally conducting. Such properties are similar to those of carbon nanotube composites, which isn't surprising as a sheet of graphene is basically an unrolled nanotube. But the graphene-based composites produced by Ruoff's chemical method are cheaper and more reliable to manufacture, so could be perfect for things that need to be lightweight, strong and conducting, such as aircraft fuselages.

Although graphene's first foray into industry is likely to be as a composite, several groups are working on unlocking the potential of the isolated sheets. For one thing, the sheets' elec-

tronic properties make graphene a candidate to replace silicon in a fresh era of microchip electronics. "Graphene is quite different from conventional semiconductors such as silicon," explains Philip Kim, a physicist from Columbia University in New York. Electrons move through silicon in a series of collisions; these generate heat and limit the speed and size of silicon transistors. The density of components on silicon chips has doubled every 18 months or so since the 1960s, a trend known as 'Moore's law' after Intel's co-founder Gordon Moore predicted it in 1965. But many believe silicon chips will soon reach their limit.

In graphene, however, electrons shoot along with minimal resistance. This, says Kim, may allow for low-power, faster-switching transistors. "But it's difficult to turn graphene off, which will make it hard to use in an electrical switch," he notes. "Really, we need new types of electronic architecture to make best use of graphene."

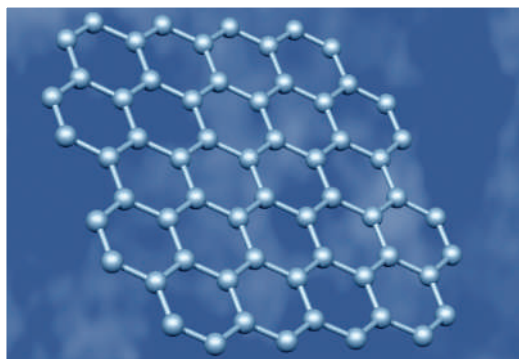
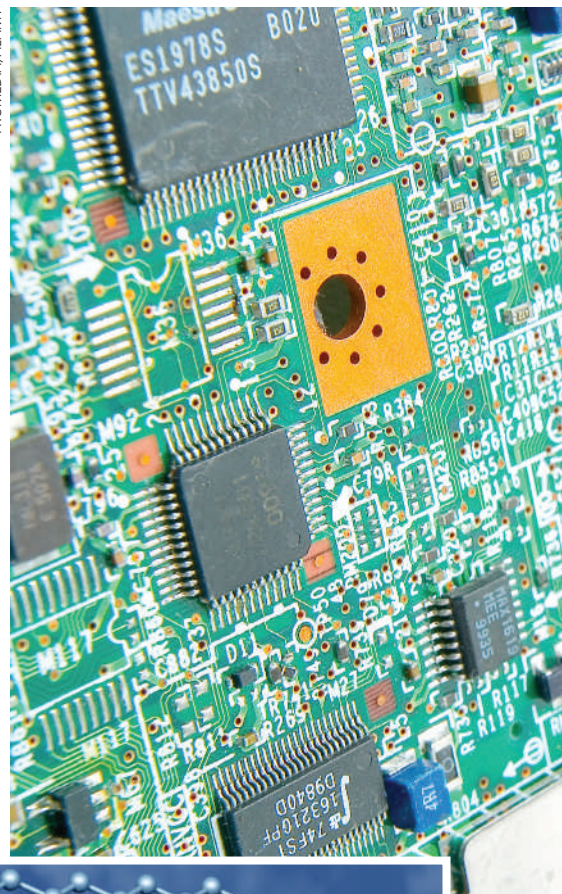
Trumping silicon

Also enthusiastic about graphene's potential in computing is Walt de Heer, of Georgia Tech's School of Physics in Atlanta. His team, with French collaborators at the National Centre of Scientific Research in Grenoble, is attempting to etch electronic structures into layers of graphene wafers grown on silicon carbide. He points out that carbon nanotubes would give the same electronic advantages as graphene sheets but that it is difficult to produce them in large amounts and organize them on a chip, and that high electrical resistances occur when nanotubes are connected with metal wires in an electronic circuit.

"No one sees nanotubes as a serious avenue," says de Heer. In contrast, flat graphene wafers are easily etched using conventional lithography techniques. De Heer also hopes that connecting wires won't be necessary. He envisages future electronic circuits made up of continuous graphene sheets, saying that cutting graphene into ribbons of different widths controls its conducting properties.

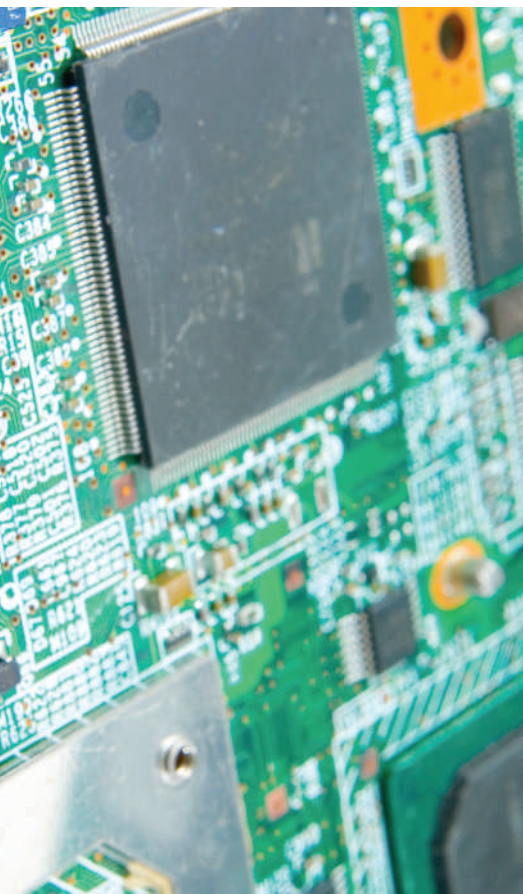
De Heer, who along with his team has

PROMEDIA/ALAMY



attracted funding from Intel, is confident in graphene's electronic potential, although he cheerfully predicts that the revolution won't come just yet. "Industry has no reason to switch over from silicon until the end of Moore's law in 10 or 15 years," he says. Geim is more wary. "Electronic applications require reliable large-scale production of graphene," he says. "At the moment the quality is rather mediocre."

Geim is more excited about graphene's usefulness as a research tool. The carbon lattice in the sheets is extremely regular, giving the electrons within the lattice unusual properties. "The electrons move collectively in a manner that mimics particle behaviour at close to light speed," explains Geim. He and other teams are using graphene to investigate quantum



Electrons pass through graphene (left) with less resistance than through silicon, making the carbon sheet a good candidate for future chips.

mechanical effects previously thought to occur only in dense plasmas around black holes and neutron stars, or in powerful particle accelerators. One example is the Klein paradox, in which fast-moving particles pass straight through a seemingly impenetrable barrier.

Meanwhile Ruoff believes his composites might yet trump other applications. By chemically modifying the sheets within the polymer matrix and studying the resulting properties, he hopes to usher in a new class of graphene-based materials. He suggests that chemically tuned composites could be used as electrically conductive plastics in solar panels, for example, or to dissipate excess heat within computer parts: "The technology wizards will take this in various directions."

One thing researchers agree on is that we're likely to hear a lot more about graphene. "Nothing is a sure deal," says de Heer. "But nothing is an insurmountable problem. All the lights are still go." ■
Richard Van Noorden

It's legal: Italian researchers defend their work with embryonic stem cells

Embryonic stem-cell researchers in Italy have reacted strongly to comments made by a Catholic cardinal earlier this month that anyone involved in destroying human embryos — including scientists deriving stem cells for research — should be excommunicated.

Concerned at the press coverage that followed the cardinal's statements, which implied that human embryonic stem-cell research is illegal in Italy, scientists from six different groups held a conference in Rome on 14 July to defend and explain their work.

The organizers included Carlo Flamigni of the Institute of Clinical Obstetrics and Gynaecology in Bologna, Maurizio Mori, a bioethicist at the University of Turin, and Elena Cattaneo, a stem-cell researcher at the University of Milan.

They also sent an open letter to Italian prime minister Romano Prodi stating that their work is legal under Italian law, and asking the government to actively promote stem-cell work and guarantee the freedom to carry out such research. "Freedom of scientific research is a principle enshrined in our Constitution. We would like reality to reflect that," the letter says.

Cardinal Alfonso López Trujillo ignited controversy when he became the first top Vatican official to publicly support excommunication of those involved in destroying embryos to derive stem cells for research. The comments by Trujillo, who heads the Vatican's Pontifical Council

for the Family, were published in Italy's leading Catholic magazine *Famiglia Cristiana* on 2 July.

The Catholic Church opposes the destruction of embryos for any purpose, as it believes that life begins at conception. The researchers behind the Italian protest say they are worried that such comments could affect political support for embryonic stem-cell



Alfonso López Trujillo wants researchers excommunicated.

research in their strongly Catholic country.

Italy already has some of the world's most restrictive laws regarding embryo research, and does not allow embryos to be created or destroyed for research purposes. However, researchers are allowed to work on imported embryonic stem-cell lines.

Cattaneo, one of the organizers of the Rome conference and a Catholic herself, says she fears further restrictions and funding cuts for the already small number of Italian groups working on embryonic stem cells. "We will remain dependent on science done abroad," she says. "In the future, I'm not sure this will entitle our country to benefit from such research."

Luca Gianaroli, scientific director of the SISMER reproductive-medicine unit in Bologna, agrees that the Vatican has a powerful influence over Italian politicians. Italy's current stem-cell law, approved by parliament in 2004, was what the Vatican wanted, he says. "It was drafted by the Church." But he doesn't believe Trujillo's comments will trigger further cuts: "It's like adding water to an already full glass."

Trujillo may also have hoped to influence politicians beyond Italy — his comments have come as politicians in both the United States and the European Union debate the use of public funds for embryonic stem-cell research. A blocking minority of countries threatens to overturn the European Parliament's decision to include such work in its latest round of research funding. And as *Nature* went to press, the US Senate was preparing to vote on a bill that would loosen restrictions on federal funding for embryonic stem-cell research.

How much influence the statement will have is likely to depend on whether Pope Benedict XVI, whose words carry far more power than those of a cardinal, feels the same way.

Since Trujillo's comments, the Vatican has remained silent on the issue. It has not voiced support for Trujillo's stand, but — perhaps more importantly — nor has it said that Trujillo was speaking personally and not for the Church leadership. ■
Jacopo Pasotti and Ned Stafford

AP



JAVA HIT BY TSUNAMI AFTER EARLY WARNING

An alert was issued minutes before the wave struck.

www.nature.com/news

AFP/GETTY

ON THE RECORD

“Even on the most solemn occasions I got away without wearing socks and hid that lack of civilization in high boots.”

A newly released letter by Albert Einstein reveals his sartorial secrets.

“There were meat-eating kangaroos with long fangs, and galloping kangaroos with long forearms, which could not hop.”

Palaeontologist Mike Archer describes some of the 20 extinct species found by his team on a fossil dig in northwest Queensland, Australia.

Sources: *New Scientist*, BBC

SCORECARD



Chinese music

Officials in China are asking the public to pick 30 songs to be broadcast from the country's first lunar probe, scheduled to launch next year.



Opera singing

Scholars in Bologna dig up the body of the world's most famous castrato singer, Farinelli, to study the physical attributes that gave him his renowned voice.



Fish stocks

Overfishing has led to a 'jellyfish explosion' in the waters off Namibia, where the biomass of the gelatinous pests is three times that of fish.

NUMBER CRUNCH

The Worldwatch Institute has mixed news about global energy trends.

19% was the amount the production of ethanol for fuel went up by between 2004 and 2005.

24% was how much global wind-power capacity increased last year.

80% of the world's energy was derived from fossil fuels in 2005.

14.6 °C was the average global temperature in 2005 — the warmest year on record.

Source: *The Worldwatch Institute*

Concerns grow over secrecy of bubble-fusion inquiry

As a way to resolve a scientific dispute, it was always likely to be fraught. In March 2005, nuclear engineer Rusi Taleyarkhan of Purdue University in West Lafayette, Indiana — known for his controversial claims to have achieved 'bubble fusion' — formally joined forces with one of his most prominent critics, physicist Seth Putterman of the University of California, Los Angeles.

But few could have predicted that this collaboration would end in such disarray. After concerns about Taleyarkhan's work were reported in *Nature* earlier this year¹, Purdue carried out an inquiry, but has shrouded the results in confidentiality, a decision that has frustrated other researchers in the field. The findings could resolve the long-standing controversy surrounding bubble fusion, but whether they will ever be made public now seems to rest on a technicality: did \$250,000 of US taxpayers' money help fund the disputed work?

The story began in 2002, when Taleyarkhan reported fusion in collapsing bubbles within a liquid, an effect also called sonofusion². His work made headlines worldwide: if the effect could be harnessed, it would promise almost unlimited energy. But others in the field were not convinced.

In the hope of settling the resulting argument, in March 2005 DARPA, the Pentagon's research agency, paid Taleyarkhan and his crit-

ics to work together to replicate the bubble-fusion experiment. Putterman was principal investigator on the \$812,000 grant and so has to account for all the expenditure. Taleyarkhan was allocated \$318,000 of the grant, and Ken Suslick of the University of Illinois at Urbana-Champaign was given \$145,000.

It was clear things weren't going well when concerns about the validity of Taleyarkhan's bubble-fusion work were reported on 8 March¹. These included an analysis by Putterman's postdoc, Brian Naranjo³, showing that the neutrons described in Taleyarkhan's latest paper, published in *Physical Review Letters* (PRL) in January⁴, came not from fusion as claimed but from the radioactive decay of standard lab material.

Purdue's provost Sally Mason called the reports "very serious" and announced a three-month review, the outcome of which she promised to publish. But once the review was complete, Purdue said the results and any future steps would be kept internal and confidential. On completion of a review, Purdue's stated policy is either to close the matter or to proceed to a fully fledged misconduct investigation. The university won't say which has occurred.

Many in the field are disappointed by the lack of information, including Putterman. He is sure that money from the shared DARPA grant was used for the work Purdue reviewed, so he is particularly keen to know whether an

WHERE DID THE MONEY GO?

The DARPA grant awarded to Seth Putterman and Rusi Taleyarkhan for work on bubble fusion began in March 2005, and Taleyarkhan submitted a paper to *Physical Review Letters* (PRL) that September⁴. Taleyarkhan insists no DARPA money was used for that work, but after checking accounts at Purdue University, where Taleyarkhan is based, Putterman believes otherwise.

- According to Purdue's accounts, the DARPA grant was billed for one-third of Taleyarkhan's salary from March until May 2005, all of it from June to August, and one-fifth of it from September to December.

- At least \$25,000 of the grant money was transferred to Taleyarkhan's former collaborator JaeSeon Cho at Oak Ridge National Laboratory in Tennessee. The PRL paper

thanks Cho for his "in-depth advice and ongoing technical assistance and cross-checks".

- Taleyarkhan's postdoc, Yiban Xu, was a co-author on the PRL paper. The DARPA grant paid all of his salary for March and April 2005, and at least half of it from May until December. (Xu's salary was originally billed at 100% for part of this latter period, but a partial refund was made in March 2006.)

- The experiment described in the PRL paper is the same as the one Taleyarkhan demonstrated to his DARPA programme manager, William Coblentz, at a meeting on 1 March 2006 to assess progress of the DARPA-funded work.

- None of Taleyarkhan's other grants, according to a list provided by Purdue University, includes the word 'sonofusion' in the title.



Rusi Taleyarkhan denies that federal dollars funded certain aspects of his research.

investigation is under way. "As a principal investigator, do I have any right to know what's happening on my project?"

In cases where federal money is involved, there is a responsibility to tell the government and the taxpayer how it was spent — and any misuse of federal dollars can have serious consequences for the researchers involved. If DARPA money was used for any of the disputed work, Purdue would be required to notify the agency of any investigation, and share information relating to it. Such communications could eventually be made public under the US Freedom of Information Act.

Taleyarkhan did not acknowledge the DARPA grant in the January *PRL* paper. But when *Nature* asked Putterman to confirm no DARPA money was used, he requested the relevant accounts from Purdue. Putterman now says: "I've reviewed the books, and I am confident that the paper relied on federal money that was not acknowledged." For example, Taleyarkhan and his colleagues claimed DARPA salaries in the run up to submission of the *PRL* paper (see 'Where did the money go?').

Taleyarkhan has declined to communicate with *Nature* directly. But he said through a third party — Brian Josephson at the University of Cambridge, UK — that Putterman's interpretation of how the work was funded is "off-base and wrong". Josephson also provided part of an e-mail in which Taleyarkhan strongly denies using the DARPA grant on the disputed work. Taleyarkhan says the experiments were completed by May 2005, several

months before the paper was submitted, and that start-up funding from Purdue paid for them. (The university told *Nature* that this funding totalled \$58,607.) He adds that he and the others involved worked on the project outside the normal eight-hour day.

In the e-mail, Taleyarkhan also says that the bubble fusion described in the *PRL* paper is different from that reported in his previous papers, on which he has warmly acknowledged DARPA funding. He does not give details of what the \$251,044 of DARPA money he spent was used for, if not the disputed work.

Taleyarkhan's explanation may make little difference if the case is investigated. "If any part of salary is allocated to a grant awarded by a federal agency, then federal funding is involved," says Mark Frankel, director of the Scientific Freedom, Responsibility and Law programme at the American Association for the Advancement of Science in Washington DC. *Nature* has confirmed this general interpretation with an investigator at a federal funding agency.

Purdue says it has not queried Taleyarkhan's assertion that no federal money was used. "The authors of the paper are the best source of information on the source of support for research," says a spokesman. "We have no reason to question the source of support stated." ■

Eugenie Samuel Reich

1. Reich, E. S. *Nature* doi:10.1038/news060306-1 (2006).
2. Taleyarkhan, R. P. et al. *Science* **295**, 1868-1873 (2002).
3. Naranjo, B. preprint at www.arxiv.org/abs/physics/0603060 (2006).
4. Taleyarkhan, R. P. et al. *Phys. Rev. Lett.* **96**, 034301 (2006).

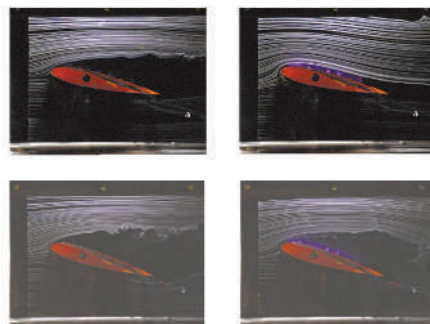
"As a principal investigator, do I have any right to know what's happening on my project?"

From aircraft engineer to FBI suspect

On his return from a trip to China on 26 May, J. Reece Roth, an emeritus professor of electrical engineering at the University of Tennessee in Knoxville, was greeted at the airport by agents from the Customs Service and the Federal Bureau of Investigation (FBI). The authorities photocopied the contents of his wallet and seized his laptop, he says. Earlier that day, they had searched his office and laboratory, and carried away hard drives and notebooks.

As *Nature* went to press, Roth had yet to be charged with a crime. "I'm still not sure what I'm being accused of," he says. "This is an Orwellian experience." Search warrants seen by *Nature* for Roth's office, laboratory and computer indicate that he is suspected of violating the Armed Export Control Act, a law that prohibits the transfer of military technologies to foreign countries or nationals.

Roth came to the University of Tennessee in 1978, where he developed a technique for creating a type of ionized gas, or plasma, in air at room temperature. Since 1994, he has been experimenting with using the plasma to control the flight of aircraft. By ionizing air as it travelled around a wing's surface, Roth created a plasma that could reduce drag dramatically (pictured above); the technique could allow air-planes to take off and land at steeper angles and on shorter runways.



J. Reece Roth's (right, inset) work with a Chinese student, to develop a drag-reducing technology for unmanned planes, has sparked an inquiry.

In 2000, Roth's work won him a three-year grant of about \$500,000 from the Air Force. The technology has also been licensed in part to Atmospheric Glow Technologies (AGT), a small Knoxville firm. In June 2005, AGT received a \$750,000 Air Force contract to develop the plasma for use in unmanned aerial vehicles. Then AGT gave Roth a subcontract to further develop his ideas.

Foreign affairs

Roth arranged to work with a Chinese graduate student who had helped him with earlier research, and contends that officials at the



University, AGT executives and Air Force administrators were all aware of the collaboration. "Everybody knew that a Chinese student would be involved, and nobody raised a red flag," he says. But when, in early May, a

Dam project threatens living fossil

We are about to lose a key piece of our evolutionary history, warn biologists. They are campaigning to save the Australian lungfish, which they fear could be sent extinct by an enormous dam planned for southeastern Queensland.

The hefty, muddy-brown fish (*Neoceratodus forsteri*) is thought to have survived virtually unchanged for at least 100 million years, making it one of the oldest known vertebrate species around and earning it the moniker of 'living fossil'. It is also one of the closest living relatives of the ancestral fish that crawled on to land and

eventually gave rise to all land vertebrates, including humans. Being able to study the species is important for understanding how that transition took place.

The lungfish is now largely confined to two river systems in Queensland — among the only places that provide the shallow, running and weedy water in which the fish likes to spawn. A dam in one of these, the Burnett river, was completed last year in order to supply water to the drought-stricken region. The area has the fastest growing population in the country, and delivering water to the

inhabitants is likely to be a huge problem in the future. But lungfish researchers say that by flooding or drying them out, the dam will eventually destroy nearly half of the lungfish spawning areas.

On 5 July, Queensland Premier Peter Beattie announced a decision to dam the second river, the Mary. Partly because the Australian lungfish is listed as a threatened species, the dam must pass a federal environmental-impact assessment before the project can proceed. But lungfish supporters believe the second dam could be enough to drive the species to extinction.

The latest decision prompted lungfish expert Jean Joss at Macquarie University in Sydney to step up a campaign to block the dam and persuade the federal government to intervene.

Joss has asked colleagues to e-mail Beattie and federal environment minister Ian Campbell to tell them about the scientific importance of the fish — so far more than 100 scientists have responded to her call. "It would be a calamitous and irreplaceable loss if this animal went extinct," says Per Ahlberg of Uppsala University, Sweden, who collaborates with Joss and



EUROSCIENCE OPEN FORUM

Read diary reports from the European science fair.
<http://blogs.nature.com/news/blog>



USAF/GETTY IMAGES

travelled to the Research Institute of Tsinghua University in Shenzhen and Fudan University in Shanghai in May, to give lectures and assist the translation of a textbook he had authored. Roth says he discussed the plasma technology generally, but didn't mention specific work being carried out by AGT. Everything in his lectures is available through the openly accessible literature, he claims.

A price to pay

FBI officials declined to comment on the ongoing investigation but Tom Reddoch, director of AGT, confirms that the company has been cooperating with the authorities. Sue Murphy, a spokeswoman for the Air Force Research Laboratory at Wright-Patterson Air Force Base in Ohio, says they too have been contacted. Murphy says that: "There's no reason to suspect any release of Air Force sensitive material."

So why is Roth being investigated? The Armed Export Control Act requires most researchers undertaking military-funded applied studies to follow a set of rules known as the International Traffic in Arms Regulations, explains Peter Lichtenbaum. Lichtenbaum previously ran export controls at the US Commerce Department, and is now a partner at Steptoe & Johnson, a law firm based in Washington DC. "With very narrow exceptions, any release of military technology to a foreign national would require a licence." Licences for cooperation with Chinese nationals are particularly difficult to get, as they require a special presidential waiver.

"It never occurred to me that a small research contract could trump the bedrock policy of non-discrimination."

It is possible that Roth's original plasma research would not have been subject to export controls, but once he was working under contract with AGT to develop specific military applications, he may have required a licence and waiver to work with his Chinese graduate student. The maximum penalty for breaching such regulations is a \$1-million fine and up to ten years in prison. But Lichtenbaum says that if charged, tried and convicted, Roth would probably face a fine of up to \$500,000.

This is not the first time that university researchers have come up against US arms export regulations. In 2003, NASA-funded astrophysicists were barred from working on Double Star, a Chinese satellite designed to explore the interaction of the solar wind with Earth's magnetosphere (see *Nature* 426, 375; 2003). But it's rare, says Robert Hardy, of the Council of Government Relations, a Washington-based organization representing research universities. "We are not aware of very many cases where a situation like this has arisen."

Roth maintains that he did not believe such restrictions would apply at his university, which encourages cooperation with foreign researchers. "It never occurred to me that a small research contract could trump the bedrock policy of non-discrimination," he says. He hopes the affair can be resolved quickly, saying the seizure of lab materials has brought his research to a halt. "This whole thing has not helped me, it has not helped the university," he says. "And it has probably not helped this country, either."

Geoff Brumfiel

newly hired contract manager at the university became aware of the situation, she apparently notified the authorities.

The concerns of the law-enforcement officials were exacerbated, it seems, when Roth



The Australian lung fish could shed light on the origin of all land vertebrates.

is helping with the campaign.

There are five other species of lungfish living in South America and Africa. But the Australian

lungfish, which can live for a century and grow 1.5 metres long, is thought to most closely resemble the last common

ancestor of land vertebrates.

Biologists say that living fish can be used for genetic and embryology studies that probe how vertebrates moved from water to land — analyses that would be impossible with preserved specimens. Joss and Ahlberg, for example, are studying the lungfish's patterns of gene activity, to try to work out how fins became limbs. "These things are amazingly important organisms in the history of the Earth," says William Bemis who studies vertebrate evolution at the University of Massachusetts, Amherst.

The Queensland government has guaranteed that the dam will include a 'fish elevator' to carry

lungfish across the dam and says that it will do whatever it takes to meet federal environmental requirements, as it did with the last dam. But Joss says that this is not enough, because the lungfish's old spawning grounds will still be destroyed. Lungfish lay very few eggs, and return to the same spawning sites year after year.

Should the campaign fail, Joss says she will petition Beattie for money to set up a lungfish breeding centre. But guaranteeing the species' survival in captivity would be tough. So far Joss is the only researcher who has managed to breed them, using two ponds, each the size of an Olympic swimming pool.
 Helen Pearson

'You can't have a mission without a camera'

A sister of the Mars Express probe has made it to Venus. And scientists have travelled to Beijing to discuss its first results. **Horst Uwe Keller**, who is presenting images from the craft's camera, spoke to *Nature* from the meeting.

I work on the Venus Monitoring Camera (VMC). One of our main objectives is to study the highest clouds, around 70 kilometres above the planet's surface.

The VMC is the only instrument newly developed for this mission, the rest were reused from Mars Express and the comet mission Rosetta. They didn't have a camera, so I stepped in and said: "You can't have a mission without a camera."

We built something very small, 1–2 kg, to replace the tiny camera on Mars Express that had watched the Beagle 2 lander as it separated.

The VMC is four cameras, one for ultraviolet light, one for visible and two for infrared. The results so far are qualitative — we haven't finished calibrating the instrument



Planet photographer:
Horst Uwe Keller.

yet. But in the sequences of ultraviolet images I presented on Monday you can see the clouds moving, showing the wind direction.

As a young postdoc in the 1970s, I remember being at a meeting where a well-known astronomer claimed that Venus's clouds were made of droplets of sulphuric acid. At the time I thought he was crazy, but now we know this is true.

What we don't know yet is the nature of the 'ultraviolet absorber' that creates dark features in our cloud images. We hope one of the spectrometers will work out what it is.

We are also working on the infrared images, which we'll check for hot spots that might be active volcanism. The surface of Venus is very young, so we know there must be some volcan-

ism, but no one has definitively seen it. If we're lucky we could also find out something about Venus's lightning using the visible-light filter.

Some of the most remarkable results at the meeting have come from VIRTIS — a spectrometer that probes the atmosphere from top to bottom. The pictures of the south pole are really spectacular: they show a double vortex created by the movement of the atmosphere between the equator and the pole.

For our team, however, the challenge is simply getting high-quality images. While the spacecraft was cruising into orbit, our camera ended up pointing at the Sun for about 50 hours. The Sun burnt its image into the photo-detector. But it's not a complete disaster: we have been able to compensate for the damage by re-measuring the sensitivity of each pixel. ■

As told to Jenny Hogan.

Police question UK science minister in loan inquiry

British science minister David Sainsbury has been questioned by police investigating links between loans to the governing Labour party and appointments to the House of Lords, it was revealed last week.

Sainsbury, who has donated £6.5 million (US\$11.8 million) to the Labour party since 2002, is the latest politician to be caught up in the 'cash for peerages' row. Officials confirmed on 13 July that the minister had spoken to detectives, although he was not cautioned or arrested.

Sainsbury, who has been in office since 1998 and is widely regarded as having performed well as minister, is a billionaire, thanks to his family's supermarket business. In April he apologized after failing to declare a £2-million loan to his party. The minister said he had confused the loan with a donation of the same amount.

For more on Sainsbury, see www.nature.com/news/sainsbury



Loan arranger: David Sainsbury failed to declare a £2-million loan to the Labour party.

Reporting of biodefence risks comes under scrutiny

A US government advisory panel has backed away from recommending that scientists receiving public funds be required to complete a checklist to gauge the biodefence risk posed by their research (see *Nature* 440, 715; 2006).

The US National Science Advisory Board for Biosecurity, which met on 13 July, said it was acting on the basis of feedback from researchers. "Scientists already feel that they're over-regulated," says Dennis Kasper, the board's chair and an expert in bacterial infections at Harvard Medical School in Boston.

The board instead approved a voluntary guide for assessing the risks of communicating sensitive biodefence research. It also released two other draft reports: a code of ethics, and criteria for identifying civilian research that could have

Smashing mission reveals what makes up a comet

When NASA's Deep Impact probe smashed into the comet Tempel 1 last July, spilling its innards, astronomers had their telescopes ready. Some of the results, published online last week (C. M. Lisse *et al. Science* doi:10.1126/science.1124694; 2006), reveal that comets are a complex mix of components.

Emission spectra captured by the Spitzer Space Telescope contain evidence for at least 16 different substances, including silicates, water ice and gas, and sulphides, providing insight into the formation of comets. The crystalline silicates, for example, are thought to have formed at high temperatures close to the Sun, and so must somehow have been



transported to more distant parts of the Solar System, where Tempel 1 was formed.

implications for bioterrorism. The reports will be open for public comment after recommendations on research oversight and increasing scientists' awareness of such dual-use work are added in the coming months.

Malaria in Africa set to get the SETI treatment

Even those lacking the billions of Bill Gates or Warren Buffett, whose philanthropy is funding efforts to tackle diseases of the developing world, can contribute to research into one of the worst killers: malaria.

Africa@home, launched on 13 July, allows volunteers to donate computing time to modelling malaria's dynamics. To participate, volunteers download a program from www.africaathome.net that simulates the transmission and health effects of the malaria parasite across Africa. The model, developed by scientists at the Swiss Tropical Institute in Basel, then runs in the background on the user's computer.

Africa@home was set up by a collaboration of academic institutions and non-governmental organizations. It is the latest in a series of scientific distributed-computing initiatives inspired by the search for alien signals, SETI@home. Other similar projects allow volunteers to simulate climate, for example, or analyse data from gravitational-wave detectors.

Plant study casts doubt on scale of methane emission

A recent surprising discovery — that plants may release large amounts of the greenhouse gas methane (*Nature* 439, 187–191; 2006) — may be much less significant on a global scale than initial lab experiments suggested, Australian scientists say.

They claim that the original researchers made an error when scaling up lab data to global vegetation cover. Because roots and woody material are unlikely to release

significant amounts of methane, only the leaf mass of trees, shrubs and grasses should have been used when calculating global methane fluxes, they argue (M. U. F. Kirschbaum *et al. Funct. Plant Biol.* 33, 521–530; 2006). They calculate that plants release between 10 million and 60 million tonnes of methane per year, substantially less than the 60 million to 240 million tonnes initially estimated.

The mechanism by which plants emit methane is still unknown, and different attempts to verify the initial findings have led to different results.

Experimental drug fails to stop prion disease

An experimental treatment for prion disease, which was at the centre of legal action between patients' families and doctors, does not seem to work, British researchers say.

In 2001, preliminary lab results suggested that pentosan polysulphate slowed the development of prion disease in mice. Families of people suffering from variant Creutzfeldt–Jakob disease went to court for access to the experimental drug (see *Nature* 426, 487; 2003). But after tracking the progress of seven patients given the drug, advisers to the Medical Research Council declared on 12 July that it does not seem to slow the brain degeneration caused by the disease.

The team did note that the patients, some of whom had been ill for several years, survived for "unusually long periods". The disease, thought to have killed 156 people in Britain since 1995, usually kills victims within a year of symptoms developing.

Correction

The News in Brief story 'Historic observatory to be preserved in luxury complex' (*Nature* 441, 800; 2006) should have said that although Edwin Hubble published observations made at Yerkes Observatory, and Subrahmanyan Chandrasekhar was affiliated with the observatory, neither of them is known to have used its 40-inch telescope.

BUSINESS

Stuck in the middle

The first foreign-based biotech company to list its shares in Japan has been caught up in bureaucracy — but its experience should help others. **Ichiko Fuyuno** reports.

Foreign companies rarely choose Japan as their base for raising money. Rules, regulations, laws and traditions make it too much hassle. But early last year, a biotechnology company set up in San Diego, California, by two Japanese doctors tried to break the mould.

The early signs were good for MediciNova: it raised US\$110 million on its initial public offering in February 2005, at a time when rivals around the world were struggling to get such flotations off the ground. With several drug candidates in the pipeline and a novel business model, based largely on taking Japanese drug candidates into the lucrative US market, the company looked set to ride high.

But these hopes soon faded. After its debut on Hercules, a stock exchange in Osaka for start-up companies, MediciNova's stock price began to tumble (see graph), and its value today is about one-third of the issue price of ¥400 (\$3.44).

Some of this poor performance is being attributed to an overall downturn in Japan's biotechnology sector. But fund managers and analysts think that it also reflects how hard it is for Japan's stock markets to accommodate firms that are based abroad — despite the government's promises to relax the tangle of regulations. "I was impressed with MediciNova's attitude," says Hisaji Agata, a managing director at JAFCO, Japan's biggest venture-capital firm. "But they didn't realize at the time that structural problems would arise later."

MediciNova was founded in 2000 as a spin-off from Tanabe Seiyaku, a midsize pharmaceutical company in Osaka that specializes in treatments for cardiovascular disease. It was led initially by Takashi Kiyozumi, a reconstructive surgeon who left in 2005, and Yuichi Iwaki, its current chief executive.

Iwaki now claims that the company has been discriminated against by parts of Japan's securities industry — including brokers and the press. He says that rules about how shares in foreign firms can be traded has made it hard for investors to buy the stock. He also argues that many information sources used by Japanese investors have been unable to include data about foreign firms — effectively shutting the company out of some investors' portfolios.

"I feel so sad that our dialogue with the market is limited," Iwaki says. "We want to break the ice to globalize Japan's securities market,



Yuichi Iwaki's company has found life tough since it was listed on the Osaka Securities Exchange (right).

but we have been having a hard time accomplishing that."

MediciNova's business model involves licensing drug candidates, primarily from Japanese companies, and then getting them through clinical trials in the United States and Europe.

The business model is attractive because Japanese drugmakers find it difficult to conduct clinical trials abroad. Iwaki, who is also affiliated to the University of Southern California, Los Angeles, has worked in the United States for 30 years, and other company board members have experience with major US biotech firms. So far, MediciNova has begun trials on six drug candidates, for treatment of diseases ranging from asthma to cancer.

Iwaki says he floated his company in Japan partly to help strengthen its ties with the business partners there who will provide its drug

candidates. But from the outset, MediciNova struggled to make its shares accessible to institutional investors. Until recently, Japanese investors who wanted to buy and sell the shares of foreign-based companies listed in Japan had to go through a subsidiary of Japan's stock exchange called Japan Securities Settlement and Custody. But institutional investors such as banks and securities firms weren't allowed to use this route. Instead, if they wanted to buy stock in Boeing, for example, they would go abroad and buy it on the New York stock exchange.

That left MediciNova and the only other foreign company to be listed exclusively in Japan — China's Xinhua Finance — in a bind. Iwaki says the problem was solved in May, when Japan's financial services agency authorized a second company, the Japan Securities Depository Center, to settle trades in foreign firms' stocks with institutional investors.

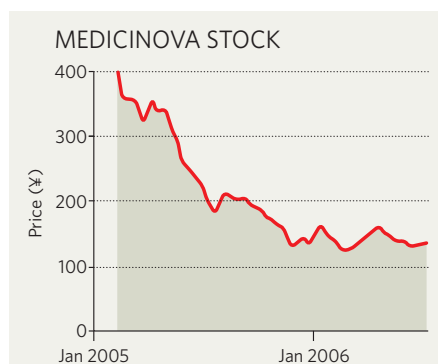
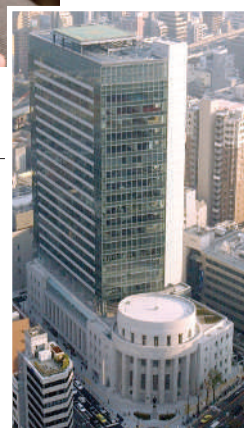
Private investors, meanwhile, faced a separate set of obstacles. The company struggled to persuade the usual newspaper listings to include its stock price. Its quarterly reports and other corporate information didn't appear on Yahoo — a favourite of private investors in Japan, as elsewhere — because the web company couldn't readily convert the information into yen, the currency used on its Japanese website. (To comply with US rules, MediciNova prepares its financial results in dollars.) These issues made individual investors wary of the stock, Iwaki claims.

In an attempt to improve communications with its shareholders, MediciNova has recently overhauled its Japanese website, and Iwaki has put up his own blog — in Japanese — to lighten up the company's image.

Some observers say that it was the company's lack of preparation for its listing on a Japanese stock market, rather than the inherent difficulties of such a listing, that got it into bother. "MediciNova was going too fast," says Chul So Moon, chief executive of the firm Cangen Biotechnologies in Bethesda, Maryland, which plans to list shares on the Tokyo stock exchange by the end of 2007. "In Japan, a step-by-step approach is important."

And according to Agata, MediciNova's experience has lessons for other foreign biotech companies that might list there. "Thanks to MediciNova, the problems of Japanese systems have come to the surface," he says. "We'll have to wait and see how the industry reacts."

■



MEDICINOV

KYODO/NEWSCOM



T. MCHUGH / SPL

Decoding our cousins

DNA extracted from bones could shed light on what happened when our ancestors crossed paths with Neanderthals. But not everyone can get the fossils out of the ground, as **Rex Dalton** learns.

In the Moula-Guercy cave, nestled in a cliff overlooking France's verdant Rhône River valley, a bent Neanderthal finger bone pokes out of the ground, as if beckoning researchers. Not so long ago, the enticement might have been ignored. With palaeoanthropologists focusing instead on the most ancient branches of the human family tree, the study of Neanderthals — our closest relatives — was something of a backwater.

But now researchers are eager to comply with the come-hither gesture. A modern-day rush is on: for Neanderthal DNA.

Across Europe, researchers are scrambling to unearth fresh bones of Neanderthals that might yield intact genetic material. Spurred by recent discoveries of DNA segments in old bones, scientists are even laying plans to sequence an entire Neanderthal genome in the next two years.

This week, scientists are gathering in Bonn, Germany, to mark the 150th anniversary of the discovery of *Homo neanderthalensis* — made in Germany's Neander Valley. During

21–26 July, experts will debate all aspects of Neanderthal life, from how they migrated across Europe to what effect climate may have had on their evolution.

At centre stage will be the future of DNA research and what it could tell us about the Neanderthals' relationship to modern humans. "Neanderthals split from the tree of human evolution about 400,000 years ago, just before the emergence of modern humans," says French archaeologist Jean-Jacques Hublin at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. "It is very important to understand in what respects genetically they are like us, and in what aspects they are different."

So far, the only closely related species with a mapped genome is the chimpanzee (*Pan troglodytes*)¹, whose ancestors diverged from those of humans about 7 million years ago. A Neanderthal genome could reveal much more about our recent evolutionary history.

The main difficulty is getting hold of genetic material. Few museum specimens of

Neanderthals have yielded DNA, and those that have could easily have been contaminated with modern human DNA. So fresh Neanderthal bones are prized possessions.

The hunt for bones is focused on the ridges and rugged valleys of southern France, where Neanderthals may have held out in their last refuge before going extinct about 28,000 years ago. They would have overlapped with *Homo sapiens* in this area for thousands of years — probably eventually losing out in the competition for resources.

Out of bounds

Alban Defleur of the Institute of Human Palaeontology in Paris hopes to return to the Moula-Guercy cave to find many more Neanderthal bones; it was here he unearthed evidence, published in a 1999 paper², that Neanderthals practised cannibalism at least 100,000 years ago. Today, the partially excavated site appears eerily frozen in time: tools are stacked about, strings marking quadrants sway from the ceiling, and a rubber bucket is

placed over the finger bone that juts from the cave floor.

Shortly after Defleur published his paper, the cultural ministry office in nearby Lyon halted digging at the cave. Officials balked at renewing Defleur's excavation permit, ostensibly because his collaborators had not filed a required sedimentology report on earlier digs. They filed it recently, saying the delay was to allow graduate students to finish their theses before details of the site were released in the report.

Some observers say in confidence that they think Defleur may have been snared in a web of nationalistic pride that overshadows French research. Co-authored with prominent US palaeoanthropologist Tim White, Defleur's article on cannibalism was published in the US journal *Science*, rather than in a French journal — a move that could have irritated some.

Other researchers have similarly fallen foul of government officials. In 1998, the same cultural ministry office that denied Defleur his permit was found by a French court to have forged documents. The forgery challenged the rights of the Chauvet cave's discoverer to data and photos; the nearby cave houses a world-famous display of rock art from this period.



Alban Defleur is still waiting for permission from government officials to re-excavate a cave in southern France that contains Neanderthal remains.

Close encounters

In many respects, France has one of the richest histories in archaeology and palaeoanthropology. The University of Bordeaux I has been a worldwide magnet, with leading researchers coming to study the country's rich Neanderthal sites.

But now a team based in Germany is generating anxiety. The Max Planck institute plans to dedicate millions of euros to the 'Neanderthal Genome Project', an international collaboration that aims to sequence ten Neanderthal genomes in the next decade³. The project is raising some eyebrows, however, among France-based archaeologists, including Catherine Hänni, of the Superior Normal University of Lyon. "Are these true collaborations?" she asks. "Or do they just want the bone for DNA?"

Hublin, who used to work as an official at the main French research agency, the National Centre for Scientific Research (CNRS), understands the concerns. "The problem is a little like oil in an Arab country," he says. "I don't want to be too critical. But France is a closed world — at times not caring

about publishing in well-rated journals."

Meanwhile, Hublin continues to push on. As part of a large team involving principal investigators from France and other countries, he is working on a hillside outside the village of Jonzac, an hour's drive inland from Bordeaux in southwestern France.

A century ago near Jonzac, miners cutting a road to an underground limestone quarry unearthed Neanderthal artefacts. But these were left largely unexplored until about ten years ago, when French geologist Emile Marchais did some preliminary excavations. Later, in 2004, researchers from Bordeaux and the Max Planck institute began a more comprehensive excavation.

When the team arrived in June to open this summer's digging season, fields of red poppies bloomed around them. Nearby, workmen prepared to erect a barn-like steel structure to shelter the expanding excavations — an indication of the project's permanence. As tarpaulin was pulled away, sending spiders scrambling, an array of artefacts was revealed — including spear points, hand axes, scrapers and animal bones. Neanderthal material is so

"France is a closed world — at times not caring about publishing in well-rated journals." — Jean-Jacques Hublin

thickly layered here that a customized computer program is needed to chart its recovery.

The artefacts are from the Mousterian group of Neanderthals that lived from about 250,000 to 35,000 years ago. The Mousterian survived in Europe at least until the first group of *H. sapiens* — the Aurignacian — appeared, some 40,000 years ago. In some Jonzac sediments dated to about 36,000 years ago, tools from the Mousterian and the Aurignacian lie close together.

Little to go on

Archaeologists Shannon McPherron and Marie Soressi, both at the Max Planck institute, take detailed notes about where artefacts from these cultures merge. They will examine the layers, to see whether they have been shifted by a flood or by the collapse of a wall of rock. If so, determining what happened at this cultural interface will be doubly hard. Archaeologists prefer ordered sediments, which are identified and dated by the type of chipped stone tool or by the fossils of extinct animals found in them.

Shifting sediment layers also complicate the search for bone or teeth for DNA analysis. *Homo* remains found in jumbled sediments must be analysed with particular care to delineate Neanderthals from modern man. Ideally, McPherron notes, he would like to find Neanderthal and *H. sapiens* specimens embedded in well-ordered sediments, showing they lived together. Such a find could provide ideal opportunities for genetic analysis — possibly showing that the two species interbred, an idea that has been long debated but never proven.

Just definitively showing that Neanderthal and modern humans coexisted at the same location has been extremely difficult. "We really, really need to understand the geology of the site to make assertions about specimens," says McPherron.

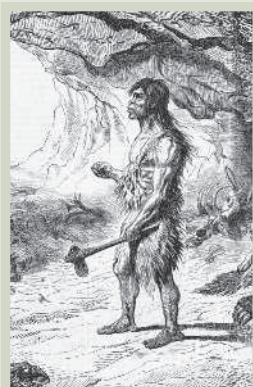
Mysteriously, given the abundance of artefacts at Jonzac, Neanderthal fossils have been scarce there, and at the team's other French site — Roc de Marsal, 200 kilometres to the east. In two years, three teeth have been found at Roc de Marsal, and one tooth at Jonzac. The team was able to extract some collagen, or connective tissue, that could hold DNA. But McPherron says there is so little collagen that they are opting not to perform DNA analysis for fear of using it up. Instead, the team plans to test for stable isotopes such as carbon and nitrogen, which could offer glimpses into the diet and lifestyle of the Neanderthals.

Swante Pääbo, a Swede who directs the Max

A. DEFLEUR

THE CHANGING FACE OF NEANDERTHALS

With only bones as a guide, artists have struggled for a century and a half to depict Neanderthals as they really lived. A few attempts over the years:



1873
This illustration from *Harper's Weekly* was published less than 20 years after the species was discovered in Germany's Neander Valley.



1929
This model, one of several displayed at the Field Museum in Chicago, presented one of the first three-dimensional views of the creature.



1950
Neanderthals are shown interacting with each other as a family in this painting from the Natural History Museum in London.



1999
A Neanderthal child from Gibraltar, perhaps as young as three at death, got a computer-generated makeover.



2006
The original Neanderthal specimen, which was discovered in 1856, comes to life in this reconstruction from the Neanderthal Museum in Germany.

Planck institute's research into palaeogenetics, is aching for more material for DNA analysis. In 1997, he published mitochondrial DNA sequences from Neanderthal remains⁴. But although mitochondrial DNA is more abundant than DNA in the nucleus of a cell, nuclear DNA provides a far more complete picture.

In May, Pääbo reported at a meeting at Cold Spring Harbor Laboratory in New York that his team had recovered and sequenced about 1 million base pairs of nuclear DNA from Neanderthal bones from Croatia. The work could provide clues about when certain disease-causing genes arose in evolution, and perhaps pin down the source of traits such as hair or skin colour.

This week in Bonn, Max Planck officials will unveil the institute's plans to sequence an entire Neanderthal genome. The project will extract the nuclear DNA from bones or teeth from both the first Neanderthal specimen discovered, from Germany, and the Croatia material. "Absolutely, I believe we can do this now," says Pääbo. But his team wants to develop new technologies to meet its eventual goal of sequencing the genomes of ten separate Neanderthals.

Search of a lifetime

The DNA bug has also bitten Defleur, who is seeking to collaborate with one of the gene-hunting teams. His group has already found dozens of fragments of Neanderthal bones at Moula-Guercy, representing at least six individuals, both juvenile and adult. "I think there are many more here," says Defleur. White, a palaeoanthropologist at the University of California at Berkeley, has excavated with Defleur from France to Ethiopia and says: "Defleur's high-quality excavation and recovery at Moula-Guercy rivals that of modern forensics."

A key advantage of the Moula-Guercy site

is that the bones are well preserved — because of the cannibalism that removes the flesh and thus destructive bacteria. Long bones have been cracked open for the marrow, which may increase the likelihood that DNA will survive, says White, an authority on cannibalistic practices⁵.

White suspects this may be why Pääbo was successful in sequencing DNA from the Croatia samples: because they were from a site, called Vindija, where cannibalism was practised.

Almost monthly, palaeogeneticists are pushing back the clock on the age of DNA extracted from Neanderthals and associated

**"Are these true collaborations?
Or do they just want the bone for
the DNA?" — Catherine Hänni**

mammals. The specimens from Vindija are roughly 38,000 years old, making them the oldest nuclear DNA yet extracted. Last month, Hänni reported the isolation of the oldest known Neanderthal mitochondrial DNA, at 100,000 years old, from the Scladina cave in Belgium⁶. And earlier this month, a Spanish-led team reported isolating 400,000-year-old mitochondrial DNA from a cave bear (*Ursus deningeri*) in the Atapuerca cavern complex in northern Spain; the same cave complex has produced nearly three dozen early Neanderthal specimens⁷.

These older specimens produce only very short genetic segments. There were 123 base pairs isolated from the Belgium site. Only 41 base pairs were found in the cave bear studies, led by Cristina Valdiosera of the Complutense University of Madrid. This is a far cry from the 3 billion base pairs that make up a *H. sapiens* or, presumably, a Neanderthal genome.

But the advances in DNA studies make Defleur think his time for finding more fossils has finally come. For more than 20 years, he has searched the French countryside for Neanderthal caves. In the early days, he worked from geological maps, checking certain limestone formations for caves. Elevations — that would have been high enough to provide protection, but not so high as to be hard to access when carrying game — were matched to the formations. Using these techniques, Defleur found new caves, a number of which held remains of Neanderthal life, and re-excavated shelters that had been left only partially examined.

Near the Chauvet cave, a rock shelter containing many lithics, called Ranc de l'Arc, is one such place probed by Defleur. "I would very much like a major excavation here," Defleur says. And there are other newly discovered caves, whose locations remain secret. These are likely to be rich sources of artefacts; some even have lithics and pottery on the cave floor.

For each journey to these dwellings, Defleur observes a moment of silence, seeming to commune with France's earliest residents. "These are places of spirit," he says. By next summer, he hopes, the government bureaucracy will have the spirit to allow him more Neanderthal discoveries.

Rex Dalton is a West Coast correspondent for Nature.

1. The Chimpanzee Sequencing and Analysis Consortium *Nature* **437**, 69–87 (2005).
2. Defleur, A., White, T., Valensi, P., Slimak, L. & Crégut-Bonnouire, E. *Science* **286**, 128–131 (1999).
3. Dalton, R. *Nature* **441**, 260–261 (2006).
4. Krings, M. et al. *Cell* **90**, 19–30 (1997).
5. White, T. *Prehistoric Cannibalism* (Princeton Univ. Press, 1992).
6. Orlando, L. et al. *Curr. Biol.* **16**, 400–402 (2006).
7. Valdiosera, C. et al. *Biol. Lett.* doi:10.1098/rsbl.2006.0515 (2006).



Arsenic-laden dirt is being excavated from this industrial site. But how do we judge a safe level for toxins?

M. DERER/AP

Safe and sound?

The White House is trying to reform environmental and health regulation across the board. But it is doing so very quietly. **Colin Macilwain** takes a look behind the scenes.

Bush declares war on environment, read an unflattering CNN headline in March 2001. It was early in the administration of President George W. Bush, and the White House had just upwardly revised how much arsenic it would allow in drinking water. Never mind that they had stepped into a trap laid by former president Bill Clinton, who sharply lowered the acceptable level of arsenic just before leaving office. The direct approach to changing the standards didn't play well, and it would be the last time the administration openly sought to relax a pollution rule.

After that, the usually direct administration took a discreet approach to reforming the way in which the US federal agencies regulate everything from drinking-water standards and air quality to airport security and work safety. The approach is so discreet that few outside Washington DC even realize it exists. Its headquarters are at the Office of Information and Regulatory Affairs, at the White House Office of Management and Budget (OMB) — the low-key powerhouse that oversees the \$2.7-trillion annual spending of the federal government.

Until this February, the effort's chief engineer was John Graham, a soft-spoken, bespectacled former director of the Center for Risk Analysis at Harvard. When Graham quit the government to run the Pardee RAND science-policy graduate school in Santa Monica, California, he left behind a proposal that could

radically change how health, safety and environmental rules are drawn up. Its real effect would be to relax them, critics charge.

The OMB Bulletin on Risk Assessment landed quietly on agency officials' desks in January. The 26-page document outlines proposed changes to the way the government conducts risk assessments — the scientific studies that quantify the risks to human health, safety and the environment caused by various activities. The bulletin is merely a proposal at

"They greeted my proposal with a long silence and a blank stare, as if risk assessment was a term from a foreign language"

— John Graham

this stage and, unusually, the OMB has sent it to the National Academies for review. An academies panel chaired by John Ahearne, an engineer and former head of the Nuclear Regulatory Commission now at Duke University in Durham, North Carolina, will report on it in November. The OMB should issue its final bulletin soon afterwards. And all federal agencies will be expected to adhere to it.

"The quality of risk assessment in the federal government is uneven," explains Graham. "What we're trying to attain is a standard across the government."

But already the knives are out for the proposal. Environmental and consumer groups charge that its real aim is to weaken regulators, such as the Environmental Protection Agency (EPA) and the Food and Drug Administration. "If this is adopted in anything close to its current form," says Rena Steinzor, a regulatory-affairs specialist at the University of Maryland in Baltimore, "it will have a devastating impact on public health and safety regulations."

Congress is also getting involved. On 5 May, leading Democrats in the House of Representatives fired off a letter to the president of the National Academy of Sciences, Ralph Cicerone. In it, Tennessee's Bart Gordon, Michigan's John Dingell and other senior Democrats ask the National Academies to spell out the limitations of its own review. In effect, they warn it to refrain from filing a scientific endorsement of what they regard as a fundamentally political proposal.

"It appears impossible to provide a comprehensive answer to the questions without reaching beyond the scope of a scientific review," says the letter. It goes on to charge that the "OMB's proposed bulletin is in conflict with the approach taken in existing law" — which, the authors say, already instructs individual agencies on how to assess particular risks.

Both supporters and critics agree that the document is the crowning achievement of Graham's long march to regulatory reform. It is the fourth important element that he put in place

from his powerful position at the OMB. Three broad-ranging edicts on how agencies should, and should not, go about the production of regulations have already been issued: they covered 'information quality', cost-benefit analyses and scientific peer review. The last drew a sharp and uninvited riposte from the National Academies when it was released in 2003. The academies said it was far too prescriptive in telling scientists how to review the work of their peers; the document was subsequently watered down to accommodate the concerns.

The main scope of January's risk-assessment bulletin is twofold: it offers guidance on how government agencies should go about conducting such assessments, and broadens the set of circumstances in which they need to be done.

Scientific risk assessments are murky affairs at the best of times. The US environmental movement, in its 1970s heyday, strongly resisted the very concept; greens prefer the rival paradigm of the 'precautionary principle'. This principle holds that a regulator responsible for, say, clean water, should respond to uncertainty about the toxic effects of a given chemical by setting a limit that it holds to be safe, in advance of more precise information. The approach is anathema to regulated industries, but it has been embraced, at least in theory, by political leaders in the European Union. By contrast, in the United States, where water and air quality are more tightly regulated, risk assessment has been incorporated into many environmental laws. With the bulletin, it would be officially incorporated into all federal-agency responses to risk.

Triple bill

Graham has cited three main examples in support of the need for the bulletin. One takes issue with a 2002 EPA estimate of the 'safe' dose in drinking water of perchlorate — an ingredient in solid rocket fuel, often dumped on public land by the Department of Defense. A subsequent, 2005 National Academies study identified a much higher safe dose of the chemical than the EPA had done — leading the agency's critics to say that it should take more care before issuing such dose estimates.

The second case also concerns the EPA, and an estimate that it made of deaths caused from cardiopulmonary disease as a result of diesel fuel emissions from off-road vehicles. The EPA made its estimate, Graham argues, even as it requested millions of dollars to research the topic. It turned out that the agency's estimate was merely a central value between risks that might be much higher or much lower, depending on the outcome of the research. "When an agency produces a risk assessment with false precision," says Graham, "it engenders negative reactions from stakeholders and from the scientific community."

His third example addresses a rather different area. The Department of Agriculture decided to shut the Canadian border to cattle



Main man: John Graham has led the White House mission to change agencies' approach to risk.

trading after mad cow disease was found in Canadian cattle in May 2003. Graham says that when he raised the issue of a risk assessment with department officials, as they were considering re-opening the border, he got "a long silence and a blank stare, as if risk assessment was a term from a foreign language".

Only last week, a National Academies panel took sharp issue with yet another EPA risk assessment — a 2003 estimate of the cancer risk posed by chemicals called dioxins. The panel, chaired by David Eaton of the University of Washington in Seattle, said that the agency had failed to quantify the risks in its assessment, or to justify the main assumptions behind it.

Graham's bulletin would require government decisions to be subjected to formal risk assessment; in emergencies, this could be completed after the decision was implemented. Deciding what constitutes an emergency can be difficult, however, and critics claim the regime would stifle government action. "They won't be able to react to things that involve a rapid decision," says Jennifer Sass of the Natural Resources Defense Council, an environmental group based in New York. For instance, the EPA moved swiftly to monitor mould growth in New Orleans after Hurricane Katrina — a response that might not be possible

under the new rules, she claims.

The measures enacted by the OMB, critics say, will provide enough sawdust to clog up the wheels of government regulation for years to come. "The aim is to bog the process down, in the name of transparency," says Robert Shull, head of regulatory policy at the pressure group OMB Watch, based in Washington DC.

Graham, a former economist, says its aim is quite the opposite. "There will be some cases where risk assessments will become more expensive" as a result of the bulletin, he concedes. "But their improved quality will mean less controversy and less delay — so we can see opportunities for saving time and money."

Fans in commerce

The measure's strongest supporters are the US Chamber of Commerce and industry groups such as the American Chemistry Council, headquartered in Arlington, Virginia. "It crystallizes 15 to 20 years of research on how best to do risk assessment, on the basis of what has been said by the National Academies and others," says the chemistry council's senior toxicologist, Richard Becker. His group has endorsed the proposal and suggests that it be given more teeth by making it subject to judicial review, in which an aggrieved party can challenge the government's implementation of it.

The National Academies is accustomed to being in the middle of this kind of fight, but it is rarely asked to review a policy proposal before it goes into force. Sceptics smell a trap: the academies' unsolicited intervention greatly diluted the previous White House proposal on standardizing peer review of scientific evidence. By asking for its advice first, the OMB can hope to implement a final rule — which it will write itself — that will be difficult for future administrations to overturn.

The only time the unflappable Graham looks ruffled is when asked if his sending of the proposed bulletin to the academies for endorsement might be regarded as a ploy to muffle political criticism. "Anyone who believes that the National Academies can be used in that way doesn't understand the process," he sniffs. "Perhaps the

OMB should be given credit for allowing its work to be criticized by the scientific community. There was no legal requirement for it to do so."

Whatever the National Academies says in November, a rule is likely to be implemented that will embed risk assessment more deeply in the decision-making process. This will be John Graham's permanent legacy, subtly moving the regulatory goalposts in industry's favour, without catching the public's eye. "It all sounds very nice and sensible," says Shull. "Politically, it is much more viable than, say, trying to weaken the drinking-water standards."

Colin Macilwain writes for Nature from Edinburgh.

"The aim is to bog the process down, in the name of transparency."
— Robert Shull

New Orleans was prepared but it was overwhelmed

SIR — Your Editorial “The gathering storm” (*Nature* **441**, 549; 2006) states: “the voters of New Orleans, many still scattered across the country, managed last month to re-elect Mayor Ray Nagin, the individual most closely associated with the city’s woeful failure to prepare itself for the storm.” As an affected resident of the region, I find this statement both untrue and an obstacle to better preparedness for future disasters that will impact on this country.

Although your statement may resemble truth for those who watched the disaster on television, for those who lived through it, Mayor Nagin is associated with a city and region that was as prepared as much as reasonably possible. If not for his leadership, many more would have died and long-term recovery would be much further behind.

A *Times-Picayune* article of 27 August 2005, two days before landfall, notes that Nagin “would stick with the state’s evacuation plan” developed by state and local officials after Hurricane Ivan. The successful implementation of this plan facilitated the evacuation of more than a million people during a 48-hour period, an accomplishment considered impossible by all the experts. A US Army Corps of Engineers study, for example, concluded that 72 hours would be required to evacuate New Orleans.

Additionally, there has been much discussion of the difficulties faced by 100,000 residents without access to private transport. However, the available data indicate that approximately 65% of people without vehicles left the city before the onset of hazardous conditions.

For those who faced serious difficulties in leaving the city, local officials provided the Superdome as a shelter of last resort. The use of the Superdome has been the subject of much criticism. However, of the 72,000 people who remained in the city after the evacuation, 26,000 were protected from Katrina’s winds and floodwaters because of this shelter. Hundreds of these people would have died if the city had not provided this shelter of last resort. Contrary to the common perception, the people in the Superdome were provided with food, water and security. The nightmare only began when the Federal Emergency Management Agency was three days late in sending the buses for post-storm evacuation.

Many people in the United States believe that the lesson to be learned from Hurricane Katrina is that a catastrophe happened because the state and local officials were not prepared. However, the evidence on the ground indicates an unprecedented level of preparedness at the state and local level. Yet a catastrophe still occurred. In my opinion, the

lesson to be learned is that, even in a context of extensive preparedness, state and local officials will be overwhelmed by natural hazards of extraordinary magnitude, and that the country as a whole needs to be prepared for this outcome. As your Editorial points out, we are likely to be entering an era of more frequent and more intense hurricane and flood disasters.

Ezra Boyd

Center for the Study of the Public Health Impacts of Hurricanes, Suite 3221 CEBA Building, Louisiana State University, Baton Rouge, Louisiana 70803, USA

Cells have long experience of dealing with UVC light

SIR — In his News and Views article “Device physics: A bug-beating diode” (*Nature* **441**, 299; 2006) describing the development of UVC-emitting diodes that promise to have major practical applications, Asif Khan states: “Earth’s ozone layer completely blocks solar light of very low ultraviolet wavelengths. Biological organisms on Earth have therefore never developed a tolerance for this ‘UVC’ radiation, and artificially generated UVC light has become a useful tool in the treatment and destruction of bacteria, yeast, viruses and fungi.”

This statement is in error on many counts. Biological organisms existed on Earth before an oxygen atmosphere, approximately 2 billion years ago, and therefore before the protection afforded by the ozone layer. During this period, cells had ample time to experience UVC (260 nm wavelength). Evolutionary pressures led to the development of multiple repair and tolerance systems targeting UVC-induced cellular damage that are now found in all three biological kingdoms. (Early evidence for repair in bacteria is shown, for example, in R. F. Hill *Biochim. Biophys. Acta* **30**, 636–637; 1958.) Loss of a UVC repair system in humans results in the UV-sensitive disease xeroderma pigmentosum, first described in 1968 (J. E. Cleaver *Nature* **218**, 652; 1968).

Even after the appearance of the ozone layer, evolutionary pressure continued because of the similarity of the photochemistry of DNA exposed to UVC and to UVB (280–320 nm). Exposure to either wavelength results in similar photoproducts: covalent adducts between adjacent pyrimidines on the DNA strands (known as cyclobutane pyrimidine dimers and [6-4]-photoproducts). The difference between DNA damage at these wavelengths is mainly different relative proportions of photoproducts and different proportions of cytosine- and thymine-containing photoproducts.

Another misapprehension is that cells need exposure to a particular agent to develop resistance mechanisms to that agent. This fails to appreciate the versatility of most biological repair systems. DNA damage can involve a wide range of altered chemical structures, but is constrained by the structure of DNA itself. Repair systems consequently target specific structural alterations as well as general classes of lesions. The nucleotide excision repair system that repairs UVC damage excises a wide variety of DNA damage, even adducts formed by an agent such as cisplatin, which was unknown to biology before its human invention in recent years.

There is therefore ample evidence that repair and tolerance systems towards UVC have existed for long evolutionary periods and have great versatility. UVC light has become a useful tool in the treatment and destruction of bacteria, yeast, viruses and fungi because it is technically easy to generate high dose rates at a wavelength (254 nm) close to the absorption maximum of DNA with low-pressure mercury sources, not because cells lack tolerance or repair mechanisms.

James E. Cleaver

Department of Dermatology and Pharmaceutical Chemistry, Auerback Melanoma Laboratory, UCSF Cancer Center, University of California San Francisco, San Francisco, California 94143-0808, USA

Same colour, many different countries

SIR — Michael Chen, in Correspondence (*Nature* **442**, 132; 2006), raised a question about the colours used in a map with the News Feature “Forward planning” (*Nature* **440**, 987–989; 2006). He was concerned that showing both the People’s Republic of China and Taiwan as yellow would make people think that they comprised one country. However, all the countries shown as having proposed repositories for nuclear waste are coloured yellow, including Sweden, Finland, the United Kingdom and France. This certainly does not mean that these pairs of neighbouring countries have been united into one.

According to the World Nuclear Association (www.world-nuclear.org), Taiwan has nuclear power reactors in operation, and therefore it should not be excluded from the issue discussed in the News Feature. I can see nothing wrong with the colours used in the map. Being simple and focused, it fulfilled its purpose.

Zhen-Ling Sun

Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

COMMENTARY



The diversity of life on Earth is in rapid decline, yet society's response to this biodiversity crisis has lacked the urgency and attention it warrants. Why is this?

Diversity without representation

For policymakers, biodiversity can present more complex challenges than climate change, argue **Michel Loreau, Alfred Oteng-Yeboah** and their co-authors. So why isn't there an international panel of experts for biodiversity?

Since the 1992 United Nations Earth Summit conference in Rio de Janeiro, Brazil, biodiversity has received increasing attention from scientists, governments and the public worldwide. There is growing recognition that the diversity of life on Earth, including the variety of genes, species and ecosystems, is an irreplaceable natural heritage crucial to human well-being and sustainable development. There is also clear scientific evidence that we are on the verge of a major biodiversity crisis. Virtually all aspects of biodiversity are in steep decline and a large number of populations and species are likely to become extinct this century. Despite this evidence, biodiversity is still consistently undervalued and given inadequate weight in both private and public decisions. There is an urgent need to bridge the gap between science and policy by creating an international body of biodiversity experts.

Although protected areas have increased slightly during the past few decades, collectively they contain only a small fraction of the world's terrestrial species and ecosystems, and the situation in the oceans is even worse¹. The forces that push towards biodiversity loss globally are much stronger than the conservation gains. Habitat destruction (notably in tropical forests, inland waters and coastlines),

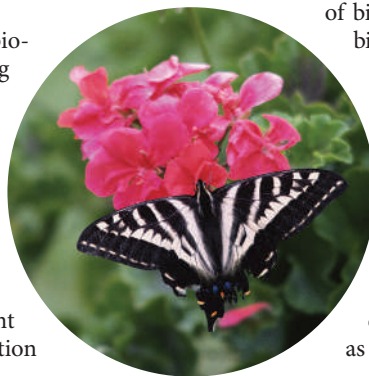
introduction of invasive species, overexploitation of biological resources (such as overfishing in the seas), pollution, and now clear signs of global climate change are major threats to biodiversity and continue unabated, driven by unsustainable growth of the world's population, production, consumption and trade.

Slow response

As a result of these forces, biodiversity loss is accelerating globally. Some 12% of all bird species, 23% of mammals, 25% of conifers, 32% of amphibians and 52% of cycads are threatened with extinction², and climate change alone might commit an additional 15 to 37% of extant species to premature extinction within the next 50 years³. Because biodiversity loss is essentially irreversible, it poses serious threats to sustainable development and the quality of life of future generations. According to the 2005 Millennium Ecosystem Assessment (MEA), two-thirds of the evaluated benefits to society from ecosystems, known as 'ecosystem

services', are currently being degraded or used unsustainably⁴.

Given the magnitude and urgency of the biodiversity crisis, why has the societal response been so slow and inadequate? A lack of awareness of the role of biodiversity in enabling the delivery of ecosystem goods and services, the failure of markets to recognize the value of biodiversity, and the fact that biodiversity is a public good are key factors. They underlie the perceived conflict between biodiversity conservation and economic development, which exists only insofar as development is not sustainable. Biodiversity is also intrinsically more complex than other environmental concerns, such as the stratospheric ozone hole or even global climate change. By definition, biodiversity is diverse: it spans several levels of biological organization (genes, species, ecosystems); it cannot be measured by simple universal indicators such as temperature and atmospheric CO₂ concentration; and its distribution and management are more local in nature.



Important ecosystem services, such as pollination, depend on biodiversity.

But another factor, which is within our power to rapidly change, is also limiting our ability to tackle the biodiversity crisis. An interesting comparison can be made between two important multilateral agreements that resulted from the Earth Summit in Rio: the Convention on Biological Diversity (CBD) and the Framework Convention on Climate Change (FCCC). The FCCC built on a strongly organized scientific community and the existing Intergovernmental Panel on Climate Change (IPCC) to inform subsequent political negotiations over climate change. In contrast, the CBD and the other international agreements concerned with biodiversity do not have the structural means to mobilize the expertise of a large scientific community to inform governments. Consequently, the scientific community often doesn't feel involved in the global political process, which tends to exacerbate the disconnect between science and policy and a general attitude of powerlessness and fatalism.

Bridge the gap

Biodiversity science needs to evolve, and is evolving, towards greater unity and integration⁵. What is lacking, however, in our view, is a mechanism akin to the IPCC that is able to bring together the expertise of the scientific community to provide, on a regular basis, validated and independent scientific information relating to biodiversity and ecosystem services, to governments, policy-makers, international conventions, non-governmental organizations and the wider public.

The four-year process that led to the MEA was a first attempt at filling the gap between science and policy. Its successes, in providing a much-needed conceptual framework and a synthesis of existing data, give us hope that the task is achievable, but it also has shortcomings for addressing biodiversity loss in the long term. First, it was a one-off effort⁶. Various projects are following up on aspects of the MEA, often at the level of a single country or region, but there is currently no mechanism for making this type of assessment global, systematic and sustained. Such a mechanism should ideally be intergovernmental, rather than non-governmental, if it is to have credibility and clout. Second, the MEA explored the consequences of ecosystem change for human well-being, but was not specifically focused on biodiversity.

The idea of establishing an international panel on biodiversity with a role similar to the IPCC has floated around for some years. But the situation is now favourable to developing and achieving it. In January 2005, the idea received political support from the French president Jacques Chirac during the

international conference 'Biodiversity: Science and Governance' held in Paris, France. It also received broad support from the 2,000 scientists, non-governmental and policy representatives from 100 countries that attended this conference, and was later enthusiastically endorsed by the 600 scientists assembled in the first DIVERSITAS Open Science Conference held in Oaxaca, Mexico, in November 2005. The French government is currently funding a consultation process to assess the need, scope and possible models for an international mechanism of scientific expertise on biodiversity (IMoSEB).

What shape would such a panel take? Although clearly its contours must emerge



Jacques Chirac speaks in support of an international panel on biodiversity.

from the ongoing consultation process, a number of features seem critical for its success. First, like the IPCC, it should have a formal link to, and be funded by, governments. This feature, which distinguishes it from previous biodiversity initiatives, would ensure that negotiations within international biodiversity conventions are based on validated scientific information and lead to action at national and global levels. However, the proposed panel should also involve other stakeholders, such as non-governmental organizations, intergovernmental agencies, conventions and DIVERSITAS — the international programme of biodiversity science.

Way forward

Second, it must be objective and independent; it should include the world's leading scientists, and its goal should be to provide rigorous, updated scientific information in support of policy decisions and actions at all levels of civil society. Just as in the IPCC, the involvement of governments and non-governmental organizations should in no way constrain the content and quality of the scientific information delivered. Third, it should be transparent and representative, in terms of opinions, disciplines and geographical regions. A strict peer-review process is required to meet these conditions.

Fourth, it should strive to generate clear,

readily accessible information about the status and trends of biodiversity, projections of future changes in biodiversity and the ecosystem services that depend on it, and options to conserve biodiversity and ecosystem services and mitigate adverse impacts of biodiversity changes. This information should allow governments, international conventions and society at large to define clear targets for action. Finally, it should build synergy with existing mechanisms and organizations. It has a unique scientific role to play, and should not attempt to duplicate existing efforts.

Building an initiative such as this requires careful examination of needs and existing mechanisms. The consultation process, supervised by an international steering committee, will last 18 months and proceed in two phases. During the first phase, a number of studies will define the need for, and goals of, an international panel on biodiversity. These studies will examine the global decision-making landscape concerned with biodiversity, analyse successes and failures of biodiversity conservation efforts at different scales, and assess existing international mechanisms that deliver scientific expertise.

In a second phase, this information will be used to articulate a set of recommendations for an international panel, which will be presented at a set of regional meetings to seek input from all sectors of society and all regions of the world.

These consultations should be seized as a unique opportunity to move biodiversity science and governance forwards and find new ways of resolving the crisis. We call upon all scientists interested in biodiversity science to get involved, and seek the participation of their government, in these consultations.

Michel Loreau is in the Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada; Alfred Oteng-Yeboah is at the Council for Scientific and Industrial Research, Accra, Ghana. This commentary is co-signed by M. T. K. Arroyo, D. Babin, R. Barbault, M. Donoghue, M. Gadgil, C. Häuser, C. Heip, A. Larigauderie, K. Ma, G. Mace, H. A. Mooney, C. Perrings, P. Raven, J. Sarukhan, P. Schei, R. J. Scholes and R. T. Watson. For additional information, e-mail to executive-secretariat@imoseb.net or consult www.imoseb.net.

1. Hendricks, I. E., Duarte, C. M. & Heip, C. H. R. *Science* **312**, 1715 (2006).
2. Baillie, J. E. M., Hilton-Taylor, C. & Stuart, S. N. (eds) 2004 *IUCN Red List of Threatened Species: A Global Species Assessment*. (IUCN, Gland, Switzerland, 2004).
3. Thomas, C. D. et al. *Nature* **427**, 145–148 (2004).
4. *Millennium Ecosystem Assessment, Ecosystems and Human Well-Being: Synthesis* (Island press, Washington DC, 2005).
5. Dirzo, R. & Loreau, M. *Science* **310**, 943 (2005).
6. Mooney, H., Cropper, A. & Reid, W. *Nature* **434**, 561–562 (2005).

D. BABIN

BOOKS & ARTS

Wanted: an Australian Volvo

Why doesn't the Lucky Country's scientific and technological strength match its natural riches?

The Australian Miracle: An Innovative Nation Revisited

by Thomas Barlow

Picador Pan Macmillan: 2006. 288 pp.
A\$25

Barry Jones

Thomas Barlow is a congenital optimist. A scientist turned journalist, he served as an adviser to Brendan Nelson, Australia's former science minister. Barlow is now chief executive of a materials company based in Sydney. He explains that his book, *The Australian Miracle*, has been written to "provide an antidote to the doom and gloom perspectives that are perpetuated about Australia's capacity to innovate and to compete in the global knowledge economy". He sees Australia as mired in "systematic complaining", a "belligerently pessimistic attitude", with "endless hand-wringing", and says it is "misguided and needlessly gloom-ridden". He has a rare gift for understatement.

I find his diagnosis puzzling. The current mindset among Canberra's ruling élite is an overwhelming, even overweening, optimism, not all of it justified. Australia's growth rates are among the highest of any developed country, and while mineral prices remain high and China's demand for raw materials remains insatiable, the country's prospects are glowing. Public discourse rarely hints at the possibility of a downturn. The mood of "Aussie! Aussie! Aussie! Oi! Oi! Oi!" is not confined to sport.

The huge salaries currently paid to Australian executives (not mentioned in the book) are not signs of depression. However, of the 2006 'Rich 200' list, published by the *Business Review Weekly* this May, only six are involved in technological services and franchising, one in developing new products. Property remains the greatest source of wealth in Australia, followed by services, retail and investment, with manufacturing well down.

I have asked scores of magnates in recent years whether they were interested in investing in biotechnology, computing, pharmaceuticals, medical technologies or new materials. The answer is invariably: "Why would I bother?" The Rich 200 are role models for the next generation of investors and entrepreneurs.

"Australia remains a lucky country in a number of ways," Barlow writes, "but luck has never been the main determinant of its success." He returns to the Lucky Country concept



AUSTRALIAN PICTURE LIBRARY

Easy money: has Australia's ability to mine resources such as iron ore hindered its development?

several times but does not mention Donald Horne's book of the same name.

Horne argued presciently in *The Lucky Country* (Penguin, 1964) that the abundance of Australia's mineral base and 'lucky' elements in its history retarded some aspects of its social, economic and technological development. People in other nations had to live by their wits or starve, but Australia always had stuff to dig up and sell, and that determined its concept of value. When Australia faced a crisis, as it did in the 1850s, 1890s, 1940s and 1950s, the luck changed — thanks to discoveries of gold, coal, oil, iron ore and natural gas, and the arrival of General MacArthur. It did not have to reinvent itself or work out new strategies.

Barlow argues that "Australia has its own extraordinary story of technological growth." Up to a point, Lord Copper. Oddly, although Australia is a heavy user of communications, computing, satellite, remote sensing, mining and military technology, it produces very little of its own, and its trade imbalance in knowledge-based products is alarming.

Barlow fails to address what I refer to as Australia's inventory problem, the conspicuous lack of internationally successful high-value-added brand-name goods and services. The

Netherlands, Sweden, Switzerland and Finland, which all have smaller populations than Australia, make products that sell internationally on reputation, rather than price. Where are the Australian equivalents of Philips, Volvo, Scania, Hasselblad, Nestlé, Roche or Nokia?

In 1986, comparing Australia and Taiwan, it would have been reasonable to assume that by now Australia would have been well ahead in information technology, given its strong education system, research history, inventiveness and position as part of the English-speaking world. In fact, Taiwan streaked ahead. Australia suffered from a failure of nerve, short-sighted thinking by institutional investors, and a lack of dynamic and compelling leadership in the computing business.

The chapter entitled "The Australian Miracle" is deeply puzzling. It is a lively and skilful précis of *Technology in Australia 1788–1988*, published by the Academy of Technological Science and Engineering. Barlow is very strong on the stump-jump plough but mentions no technological innovation later than 1900. The word miracle, suggesting divine intervention, or beyond reason, is an odd choice not justified in the text.

Barlow could have put far more emphasis

on Australia's outstanding record in medical research. He rightly praises Barry Marshall and Robin Warren for their work on *Helicobacter pylori*, which won the 2005 Nobel Prize in Physiology or Medicine. But Macfarlane Burnet, Peter Doherty and Suzanne Cory are barely mentioned, and Jack Eccles, Frank Fenner, Don Metcalf, Gus Nossal and Jacques Miller are ignored.

The most thoughtful element in Barlow's book is a brief examination of the comparative value of individual versus collaborative research; he should have expanded this argument. He expresses concern that Commonwealth and state funding of research may involve too much fanfare about multidisciplinary at the expense of research by individuals. Darwin's research and Einstein's celebrated 1905 papers would have been too small to qualify for Australian Research Council (ARC) grants, and I suspect that Burnet, a notorious loner, might have looked in vain for funding from the National Health and Medical Research Council.

The proof-reading is careless. We are treated to a quotation from Arthur Koestler twice. And Barlow dismisses Australia as a "relatively small economy", even though *The Economist* ranks Australia as number 15 in the world. The book also suffers from the lack of an index.

Barlow is right to point out the Australian public's high level of interest in science, but is dismissive, even contemptuous, about complaining academics. He fails to examine the reasons for this paradox. He should have commented on the striking fall in enrolments in the enabling sciences, physics, chemistry and mathematics, parallel to the relentless march of the business students. Research disciplines are down proportionally (but not in absolute numbers) while the packaging and marketing subjects are well up. Was this worth a sentence or two?

Barlow is unduly modest and says little of his time advising Nelson. He mentions the case of Trofim Lysenko, the Soviet geneticist who suppressed darwinism, and asserts that nothing similar could happen in Australia. I am not so sure. Australia practises its own form of soft lysenkoism, with climate scientists in CSIRO silenced if they do not produce 'agreed science', supporting the Howard government's ideological rejection of global warming. 'Public good' Cooperative Research Centres were closed down unless they worked on commercial products. The ARC's recommendations were subject to ministerial veto, and a chair and chief executive appointed for expertise in research were replaced by people experienced in working with government and interpreting, or even second-guessing, shifts in policy. ■

Barry Jones is a vice-chancellor's fellow at Melbourne University. He was Australia's science minister from 1983 to 1990. His autobiography, *A Thinking Reed*, will be published in October.

Renaissance man

Thus Spoke Galileo: The Great Scientist's Ideas and their Relevance to the Present Day
by Andrea Frova & Mariapiera Marenzana
transl. by Jim McManus

Oxford University Press: 2006. 512 pp.
\$34.50, £19.99

Giorgio Parisi

Galileo Galilei is known as one of the founders of modern science. But his works, as well as those of co-founders such as Isaac Newton, are seldom read. It is not that they are in some exotic language: Newton wrote in Latin, Galileo mainly in a limpid, straightforward Italian, in contrast to the baroque style of many of his contemporaries. As well as science, Galileo is a major figure in Italian history and literature. Italian high schools teach his dramatic life and prosecution by the Inquisition. But professional historians of science aside, few Italians have read many of his works.

The problem is that we must understand the scientific questions, knowledge and — most importantly — prejudices of the time. Changes in scientific notation make the scientific papers of even a century ago all but unintelligible to today's experts, so it is no surprise that scientists are seldom moved to read 400-year-old books of physics whose scientific context almost completely escapes them.

Thus Spoke Galileo by Andrea Frova and Mariapiera Marenzana, an attempt to bring Galileo's work to general readers, should solve this problem. The texts are introduced in their

correct and precise historical context, framed to stimulate the reader's curiosity.

The book opens with an imaginary autobiography, composed of a collage of letters and other documents written by Galileo, interpolated with writings from Frova and Marenzana in Galileo's style. Galileo tells his life story, discussing the clash with the Church that eventually led him to disown his writing. It is a portrait of a complex man, with light and shade. He was frank, but also able to adapt himself to difficult times: he knew that his forced abjuration was only a momentary defeat, and that in the long run, with the help of his students, his ideas would triumph.

Each of the other chapters deals with a specific topic. They begin with a summary of the knowledge at that time, followed by Galileo's writings on the subject. The most crucial points are put in his words; the rest of the argument is summarized. Often the chapter ends with a short historical excursus where, presenting the same ideas in modern scientific language, the authors show the impact of Galileo's ideas on subsequent science.

To understand the originality and ingenuity of a scientist we must compare his statements with those of his contemporaries. It turns out that Galileo leant on the work of other scientists more than is commonly believed: the famous example of the ship in motion is taken from Giordano Bruno, and an argument on the fall of a heavy body comes from Giovanni Battista Benedetti. Galileo's unique skill was to

Galileo in the gallery

Galileo's influence on science and culture is celebrated in art and on stage in Britain this summer.

Bertolt Brecht's play *The Life of Galileo* examines the conflict between reason and faith, a theme with current resonance. This month sees the revival of a version of this play by David Hare at the National Theatre in London. Directed by Howard Davies and starring Simon Russell Beale, it runs until the end of October.

Galileo's book *Sidereus Nuncius* (*The Starry Messenger*) lends its title to an exhibition at the Compton Verney gallery in rural Warwickshire. *The Starry Messenger: Visions of the Universe* explores how science and technology have changed the way we think about the Universe, and the role of artists in transmitting these ideas. Early scientific texts are on display alongside paintings, music and video installations by artists ranging from William Blake to Fred Tomaselli (shown here). The exhibition runs until 10 September.

► www.nationaltheatre.org.uk
► www.comptonverney.org.uk



JAMES COHAN GALLERY

weave empirical observations into a coherent picture. He also used state-of-the-art technology to build and develop instruments of scientific observation, such as the telescope and the microscope, that, through their widespread use, went on to fundamentally influence later science.

The book is remarkable for its clarity, precision and historical accuracy. Numerous drawings, figures and photographs help the

reader pick a path through the historical and scientific reconstruction.

Nowadays scientific knowledge is extraordinarily developed and university education promotes hyper-specialization. Scientists who end up working in two different fields may struggle, owing to an insufficient depth of knowledge in either one. This book reminds us that things have not been always this way: as the authors note in their introduction, the excerpts from

Galileo's writings have been chosen to offer a vision of his multiple interests in many scientific and cultural fields, "of his prodigious curiosity for all natural phenomena, of his inexhaustible capacity for posing questions to himself and searching for answers by reasoning. And also of his exultation at making discoveries."

Giorgio Parisi is in the Department of Physics, University 'La Sapienza', Piazzale Aldo Moro 2, Rome 00185, Italy.

A drink from the magic well

Asian Honey Bees: Biology, Conservation and Human Interactions

by Benjamin P. Oldroyd & Siriwat Wongsiri
Harvard University Press: 2006. 340 pp.
\$59.95, £38.95

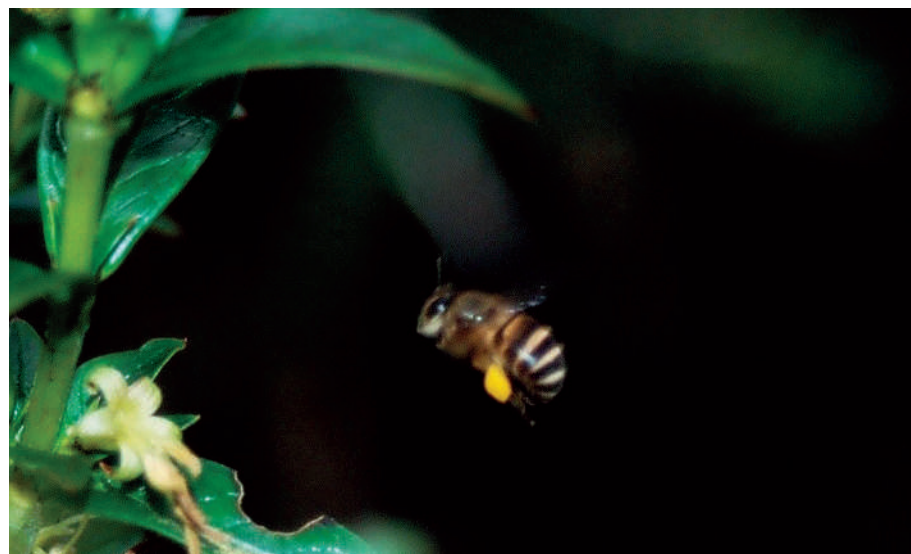
Francis L. W. Ratnieks

Karl von Frisch, who won the 1973 Nobel Prize in Physiology or Medicine for discovering the honeybee dance language, called the honeybee a magic well — no matter how much was discovered, more remained. Most research on honeybees, including von Frisch's, has been done on *Apis mellifera*. This species, the western hive bee, is native to Europe, Africa and the Middle East, and has been introduced worldwide because of its economic importance in making honey and pollinating crops. But Asia is home to eight more species of *Apis*, five of which have been recently discovered or recognized. The past ten years has seen an upsurge of research on these species, leading to some 250 scientific articles.

Four of the Asian honeybee species nest, like *A. mellifera*, in cavities within which they build multiple vertical wax combs. There are also two species each of dwarf and giant honeybees. These build a single vertical wax comb, usually under a branch or, in the case of giant bees, an overhang. All live in tropical Asia, except for the eastern hive bee, *A. cerana*, which extends as far north as Japan and Korea, and the giant mountain honeybee, *A. laboriosa*, which lives in the Himalayas.

Asian Honey Bees has a simple but effective structure. It begins with chapters on honeybees in general, the Asian species, honeybee evolution, speciation and biogeography. These provide the context in which to understand the biodiversity and evolution of honeybees and also for the following five chapters on particular topics: dance communication and foraging; reproduction; worker sterility and kin selection; nesting and defence; and parasites and predators. The final two chapters address interactions with humans and conservation.

Studies of Asian honeybees have done much to elucidate the evolution of the dance language, a research question initiated more than 50 years ago by von Frisch's student Martin Lindauer, and many other areas of honeybee



Never mind the buzz about *Apis mellifera* — *A. nuluensis* is one of eight other honeybee species in Asia.

biology. These studies have also discovered examples of nature at its most wonderful, of which two deserve a mention here. Workers of *A. cerana* defend their nest against giant hornets by encasing the hornet in a living ball of bees, which generate heat and literally cook the hornet to death before it can send a pheromone signal to recruit more hornets to the attack. Indeed, the scout hornet's pheromone alerts the bees. Colonies of giant honeybees, *A. dorsata*, often migrate 100 km. At certain times of the year the bees abandon their nesting trees, flying to better areas for a season before returning. Genetic studies show that the same colony with the same queen may return months later to the same tree.

Asian Honey Bees succeeds admirably. It is both an authoritative monograph that will satisfy experts and a highly readable book that will engage students and biologists in general. It presents Asian honeybees in the broader context of social behaviour and evolution while giving a real feeling for the bees themselves, including their interactions with humans — you can buy bee brood in many Asian food markets, something the authors have used as a way of getting samples.

The book itself is well produced, although I have one quibble. The numerous photos are

reproduced in black and white. Surely insects demand colour! But this is partly offset by a stunning cover photograph of a nest of the giant honeybee, *A. dorsata*.

Francis Ratnieks is in the Laboratory of Apiculture and Social Insects, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

MORE ON BEES

Letters from the Hive: An Intimate History of Bees, Honey, and Humankind

by Stephen Buchmann with Banning Repplier
(Bantam Books, \$14)

A book interweaving the natural history of bees with their small but vital place in world culture. The story ranges from prehistoric honey-hunting to mead-making, pollination, bees in warfare, and the medicinal properties of honey.

A Keeper of Bees: Notes on Hive and Home

by Allison Wallace (Random House, \$23.95)
A beautifully written book that peers into the hive for a hard look at honeybee behaviour, then swerves all over the place, covering the events that first sparked the author's interest in bees, the effects of pollution, bees' place in myth and nature, and the many hives the author has created over the years in various back gardens.

NEWS & VIEWS



E. BAKER

Figure 1 | Message from the deep — the flank of the Dabbahu volcano, evidence for extensive dyking below.

PLATE TECTONICS

Magma does the splits

Freysteinn Sigmundsson

A minor volcanic eruption in Ethiopia was the main visible clue to a massive injection of magma along the Afar rift last year. Such inconspicuous processes could have been crucial in early continental break-up.

Earth's surface is a mosaic of cool, strong rock plates that are moving continuously at rates of up to 12 centimetres a year. On page 291 of this issue¹, Wright *et al.* report studies of tectonic events in Ethiopia that provide an unusual insight into what happens in Earth's crust when two of these plates diverge.

Divergent plate movements cause a build-up of tensional stress that is released in a process known as rifting. This consists of periods of large-scale faulting (the slippage of adjacent rock masses at fractures in Earth's crust) and/or volcanic and magmatic activity. The relative contributions of fault movements and magmatic activity are unknown, as are how the deformation processes vary with depth and when exactly rifting episodes occur. Opportunities to observe rifting are rare: separation occurs mostly on mid-ocean ridges and in only a few places above sea level. At any one location, a rifting episode might occur only once every few centuries.

One place where rifting can be observed on dry land is along the East African and Afar rifts in Ethiopia, where the Nubian plate diverges from the Somali and Arabian plates². The quiet flow and accumulation of magma over a long period — from at least April 2004 to May

2005 — 5 kilometres below the surface of the Afar rift preceded a series of more than a hundred earthquakes in September 2005, of maximum magnitude 5.6, that were registered across the globe. Continuous earthquake activity, including a volcanic tremor, was observed on 24–26 September at the nearest seismic station in Addis Ababa, around 400 km away. A minor eruption occurred on 26 September along a 500-metre-long fissure on the flank of the Dabbahu volcano in central Ethiopia (Fig. 1), and ground fracturing occurred along a rift segment that was around 60 km long.

Wright and colleagues¹ have revealed the exact course of these events, by interferometric analysis of images acquired by the European Space Agency's Envisat satellite using the synthetic-aperture radar method. They compared images taken from the same location in space before and after the event by 'subtracting' the phase signal of one image from that of the other to form a map, accurate to a few centimetres or better, of the change in distance from the ground to the satellite. By combining observations from ascending and descending satellite passes, as well as direct measurements of offsets in radar amplitude images, the full three-dimensional crustal

deformation field could be evaluated.

The results are clear. Beneath the surface fractures, magma started flowing rapidly from magma chambers 5 km under the neighbouring Gabho and Dabbahu volcanoes. But rather than flowing vertically towards the surface of the Earth, this magma flowed horizontally tens of kilometres, forming a 60-km-long vertical 'dyke' at a depth of 2–9 km (Fig. 2, overleaf). The dyke might also have been fed by a vertical flow of magma from greater depth. These magma flows split the crust laterally by up to 8 metres, relieving centuries of tensional stresses caused by plate divergence. The minor eruption on the side of Dabbahu on 26 September was probably an 'accidental' by-product, a result of the dyke heating an existing isolated pocket of magma in the crust. The inferred average rate of magma flow throughout the process was about 2,000 cubic metres a second, comparable to rates in large volcanic eruptions. The subsurface dyke now has a volume of some 2.5 cubic kilometres, twice the volume of material blown out in the eruption of Mount St Helens in 1980.

Wright and colleagues' study¹ demonstrates that dyking may relieve the tensile stress caused by plate divergence, while allowing

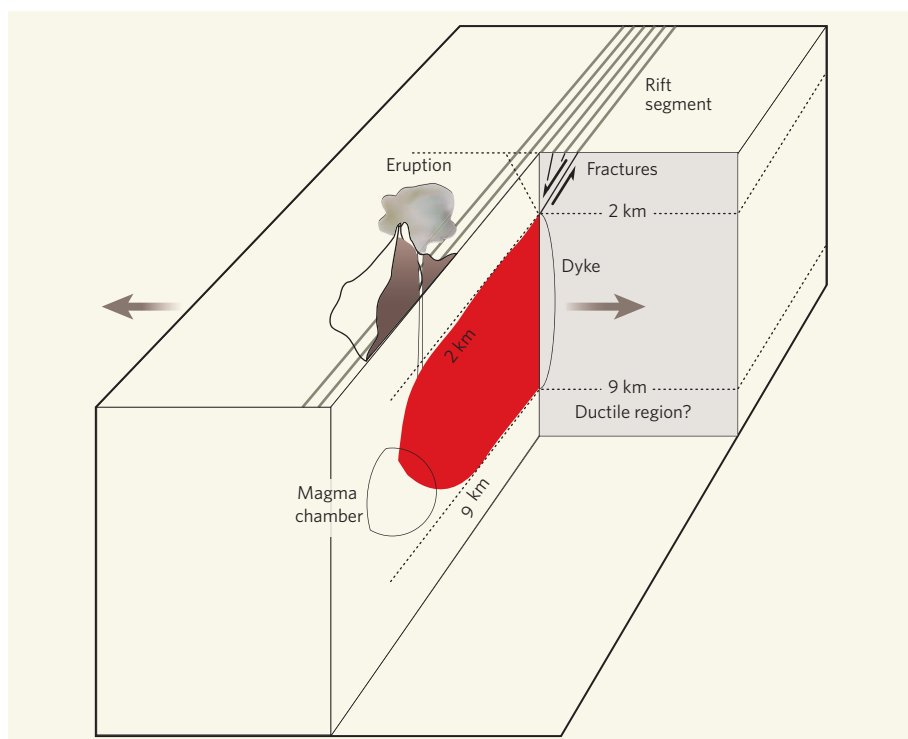


Figure 2 | Rifting in the Afar in 2005 in simplified cross-section. Movement on the fractures was mostly limited to the uppermost 2 kilometres of the crust, modelled by Wright *et al.*¹ as slip of up to 7 metres on two inclined fault planes. Farther below, magma did the splitting, forming a dyke that displaced adjacent plates by 2–4 metres in each direction. No source of co-rifting deformation is detected in a possible ductile region below about 9 km; total crustal thickness is about 20 km. The dyke was supplied by magma chambers below the Gabho and Dabbahu volcanoes, and possibly by a deeper source.

only a small amount of magma to reach the surface. Continental break-up starts with normal faulting processes and the formation of characteristic 'grabens', where land sinks down between parallel fault systems as the plates on either side diverge (the Great Rift Valley of eastern Africa, of which the Afar rift is part, is a classic example of such a sunken tract of land). Mantle rock, displaced as a graben forms, moves towards the surface, where the lower pressure causes its partial melting. The resulting generation of magma later leads to the onset of voluminous volcanic eruptions³.

But the new observations imply that many grabens and single faults, even at continental rifts, might reflect dyke-induced faulting^{4,5}, and might form and evolve through repeated events such as those that occurred in Ethiopia in 2005. An intermediate stage in continental rifting can be envisaged, when melt has begun to be generated in the mantle, but is produced in limited quantities so that most of it is captured inside the extending crust. This period will be dominated by dyking, but characterized by few eruptions.

Another argument for dyking causing early injection of magma into the crust during continental break-up is that it requires less force to dilate brittle crust by upwelling magma than by the slippage of plates in faulting. A stress of some 65 megapascals is needed to initiate slip on a normal fault at a depth of 5 km (ref. 6). In contrast, magma in contact with stretched crust will lead to hydraulic fracturing when

the excess pressure of the magma is about as large as the tensile strength of the crust. This has been inferred to be 1–6 MPa for oceanic crust in Iceland⁷. Thus, any flow of magma into the crust could cause rifting long before the crust is stretched to the degree required for a conventional faulting event.

The study of this rifting episode in Ethiopia increases our understanding of similar events that are taking place on mid-ocean ridges⁸, or where these are exposed on land in Iceland and in Djibouti on the Horn of Africa, where the

Gulf of Aden ridge meets the Afar rift. A smaller-scale rifting episode occurred in Djibouti in 1978 (refs 9, 10). The 1975–84 Krafla episode in Iceland^{4,7,11}, however, was more similar to the events of 2005 in Ethiopia in terms of its cumulative widening, the magma volume involved, and the importance of subsurface dyking during its initial four years, when there was little eruptive activity.

The deformation pattern associated with the Ethiopian events¹ suggests that there is a strong vertical variation in the mechanisms that caused them. Faulting seems to have occurred only in the shallowest crust down to a depth of around 2 km, and dyking only in the mid-crust. There is no evidence that the dyke extends below 9 km, indicating that continual ductile deformation processes must prevent the accumulation of tensile stress in this area. A lesson from Iceland is that a rifting episode might consist of many dyking events with an overall duration determined by the period of the magma inflow. More dyking and eruptions might therefore follow in Ethiopia if magma continues to flow to shallower levels.

Freysteinn Sigmundsson is at the Nordic Volcanological Center, Institute of Earth Sciences, University of Iceland, Sturlugata 7, Askja, 101 Reykjavik, Iceland.
e-mail: fs@hi.is

1. Wright, T. J. *et al.* *Nature* **442**, 291–294 (2006).
2. Sella, G. F., Dixon, T. H. & Mao, A. *J. Geophys. Res.* **107** (B4), 2081 (2002).
3. Hayward, N. J. & Ebinger, C. J. *Tectonics* **15** (2), 244–257 (1996).
4. Einarsson, P. & Brandsdóttir, B. *J. Geophys.* **47**, 160–165 (1980).
5. Keir, D., Ebinger, C. J., Stuart, G. W., Daly, E. & Ayele, A. *J. Geophys. Res.* **111**, B05314 (2006).
6. Turcotte, D. L. & Schubert, G. *Geodynamics* (Cambridge Univ. Press, 2002).
7. Sigmundsson, F. *Iceland Geodynamics: Crustal Deformation and Divergent Plate Tectonics* (Springer-Praxis, Chichester, 2006).
8. Delaney, J. R. *et al.* *Science* **281**, 222–230 (1998).
9. Abdallah, A. *et al.* *Nature* **282**, 17–23 (1979).
10. Cattin, R. *et al.* *Earth Planet. Sci. Lett.* **239**, 352–367 (2005).
11. Tryggvason, E. *Bull. Volcanol.* **47**, 47–69 (1984).

ECOLOGY

Asymmetry and stability

Robert D. Holt

Ecological communities are dauntingly complex. Nonetheless, ecologists gallantly persevere in eliciting insights about the factors that govern the behaviour and persistence of these messy, tangled webs.

The concept of symmetry is a guiding principle in much of science, ranging from fundamental physics¹, to analyses of sex-ratio evolution², to the dynamics of persistent populations³. Yet many phenomena reflect the action of strong asymmetries, including ageing⁴, the evolution of species' niches⁵, and even the grand sweep of organic and human

history^{6,7}. On page 265 of this issue, Rooney *et al.*⁸ grapple with the problem of understanding the stability of complex food webs, and argue that strong asymmetries in energy flow and interaction strengths may be key determinants of ecological stability. The word 'stability' has many meanings, but in this context it refers to the capacity of a system



Figure 1 | Top predator. The yellow garden spider (*Argiope aurantia*) is common on the fringes of ponds in northern Florida. Its prey include terrestrial insects such as grasshoppers and crickets, which are sustained by terrestrial vegetation and detritus, and aquatic insects such as caddisflies and damselflies, which when immature live in ponds and so are ultimately sustained by aquatic plant life. Several distinct flows of energy and nutrients are thus merged in the diet of this generalist predator, which is an example of the kind of top predator considered by Rooney *et al.*⁸ to be a vital link among different food-web channels.

to recover from perturbations in species abundances.

Ecological communities are among the most complex entities studied by scientists, as they are composed of thousands of species with many distinct lifestyles, interacting in a myriad of ways. Understanding the relationship between the complexity and diversity of ecological systems, and their stability and persistence, is a perennial challenge in ecology. The conventional view that 'diversity begets stability' was overturned by the seminal work of Robert May⁹ in the early 1970s. May demonstrated mathematically that increasing the diversity and connectedness of randomly assembled model food webs typically decreased their stability. So if complexity is to enhance ecosystem stability in real food webs, it must emerge from particular structural features of ecological systems.

The authors⁸ weave together a survey of empirical studies of food-web energetics with theoretical arguments to suggest that a particular pattern of energy flow found in many food webs may contribute strongly towards stability. At their base, many food webs consist of groups of species that use distinct sources of energy and nutrients. For instance, a lake food web may be sustained by both phytoplankton (which create biomass via photosynthesis) and microbial decomposers (which consume detritus washing into the lake). Phytoplankton and decomposers are very different potential resources for consumers, so organisms in the lower trophic ranks tend to be specialized. In effect, there are distinct 'channels' for inputs of energy and materials into the biota of the lake.

But as one ascends the food web, these channels begin to merge, so a predatory fish such as the lake trout might be sustained ultimately by both phytoplankton production and detritivory.

Rooney *et al.*⁸ examine published food-web data from a wide range of aquatic and terrestrial ecosystems, such as marine upwelling regions, estuaries, and agricultural and forest soils. They conclude that the pattern of channels that are distinct at the base of the food web and fuse as energy percolates up the web is widely observed. Top consumers thus integrate a broad base of alternative energy sources (Fig. 1). Crucially, the authors also observe that the different channels are not equally important, but typically exhibit strong asymmetry, with considerably more resources flowing through one channel than through others, but without complete dominance by any particular channel.

The organisms that define these distinct resource channels differ in properties such as body size, generation length and accessibility for higher consumers. These organismal traits have dynamical consequences for the food web as a whole. The authors relate a principal ecosystem concept (turnover rate, which determines how long a given individual takes on average to be replaced) to a principal community concept (interaction strength, which governs how much one species changes in abundance following a change in another's abundance). They suggest that faster channels on average reflect strong interactions, and slow channels, weaker interactions.

This in turn influences the timescale over which the channels in isolation would respond to disturbance. This difference in response

times means that a predator using both fast and slow channels enjoys a more dependable prey base than predators specialized to just one or the other. But there are more subtle implications of coupling channels that emerge from a theoretical study of a model of linked food chains. Assessing the response of the model to both small and large perturbations, Rooney *et al.*⁸ show that, in general, asymmetry in the amount of energy flowing through the two levels, and in the pattern of predator attacks, promotes the most stable food webs. They suggest that this is because the simultaneous presence of slow and fast channels prevents dramatic overshoots following large perturbations, while also permitting rapid recovery.

Thus, the asymmetric patterns that the authors document in natural food webs are also the very kinds of asymmetry that promote stability and resilience in theoretical models. There are, of course, many questions that one can pose about the generality of these results. Most of the empirical studies included in the analysis involve highly aggregated data, and the authors' model in essence describes the dynamics of just five interacting species. Including additional species and more complex assumptions about individual behaviour and interspecific interactions can lead to surprises in ecological models¹⁰, and it seems likely that the effects emphasized by the authors could be obscured by other factors in some systems.

Nonetheless, Rooney *et al.*⁸ provide an innovative empirical and theoretical contribution to this central ecological problem of identifying potentially key structural features of food webs that promote persistence. As the authors note, human impacts ranging from overharvesting to eutrophication have the effect of reducing the importance of top predators, and effectively homogenizing the energy channels, thus endangering the stability of natural ecosystems. Symmetry is elegant, and often desired in art, human affairs and the formulation of some fundamental theories in science. But in the messy, tangled web of ecological systems, a fair dose of asymmetry may be required to prevent these systems from self-destructing. ■

Robert D. Holt is in the Department of Zoology, University of Florida, Gainesville, Florida 32611-8525, USA.
e-mail: rdholt@zoo.ufl.edu

1. Lederman, M. L. & Hill, C. T. *Symmetry and the Beautiful Universe* (Prometheus, Amherst, NY, 2004).
2. Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, 1982).
3. Royama, T. *Analytical Population Dynamics* (Chapman & Hall, New York, 1992).
4. Rauser, C. L. *et al.* *Age Res. Rev.* **5**, 14–32 (2005).
5. Holt, R. D. *Evol. Ecol.* **10**, 1–11 (1996).
6. Diamond, J. *Guns, Germs, and Steel* (Norton, New York, 1997).
7. Vermeij, G. J. & Dietl, G. P. *Paleobiology* **32**, 173–178 (2005).
8. Rooney, N., McCann, K., Gellner, G. & Moore, J. C. *Nature* **442**, 265–269 (2006).
9. May, R. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).
10. Abrams, P. A. *Am. Nat.* **141**, 351–371 (1993).

MATERIALS SCIENCE

Carbon sheet solutions

Nicholas A. Kotov

When carbon fibres just won't do, but nanotubes are too expensive, where can cost-conscious materials scientists go to find a practical conductive composite? The answer could lie with graphene sheets.

Carbon nanotubes have revolutionized materials science, with their combination of exceptional mechanical strength and unique electrical properties. The media has added to the hype, by enthusing over the way this new form of carbon will transform the transport and electronics industries¹. Yet many problems still need to be solved before carbon nanotubes can be successfully incorporated into composite materials. The three biggest issues are the fact that nanotubes tend to clump together during processing, the difficulty of controlling their diameter and the way the carbon sheet is rolled, and the high cost of their production².

With all the excitement about nanotubes, an alternative option has been overlooked. Single layers of graphite — known as graphene sheets — also have excellent electrical properties, but are cheap to make and require no helicity control. Yet graphene sheets have problems of their own, to do with their poor mechanical properties. Graphite is soft and flaky, and cannot be used in load-bearing

structures. This problem could be solved by making a composite material of graphene sheets and polymers. But until recently, preparing such composites was extremely difficult, because the strong surface–surface attraction between the sheets prevents them from dispersing in a polymer solution or melt. In this respect, graphene sheets are even harder to deal with than carbon nanotubes. On page 282 of this issue³, however, Stankovich *et al.* describe a method that overcomes the difficulties of making graphene–polymer composites.

The challenge was to find a process that yielded a uniform distribution of graphene in a polymer matrix. The authors³ began by converting graphite into graphite oxide in an aqueous medium. This well-known process adds oxygen-based chemical groups to the graphite surface, and results in the bulk graphite being completely separated into single sheets. The oxygen-based chemical groups tend to have excess negative charge, so

the sheets repel each other, producing stable dispersions. To restore graphene's unique properties, the oxygen-containing groups must be removed; however, without the negative charges the sheets immediately coalesce.

Stankovich *et al.*³ struck on the idea of chemically modifying the surface of graphite oxide. They did this by treating the material with phenyl isocyanate, which adds hydrophobic chemical groups to the surface. The resulting material forms separated sheets that can be mixed with solutions of many commercial polymers in polar organic solvents. One may speculate that the hydrophobic groups attached to the graphite oxide sheets are attracted to similar groups in the polymers, so that the sheets prefer mixing with the polymer rather than stacking up with each other. The resulting composite material was an insulator. So, to restore the conductivity of the graphite, a small amount of a reducing agent was added to the composite solution. This did not make the graphene sheets coalesce, because hydrophobic groups remain attached to their surfaces, holding them within the polymer.

In this way, the authors obtained a composite with excellent structural characteristics: all the sheets were individually and uniformly distributed throughout the volume of the polymer. The composites were also easy to process using standard industrial technologies such as moulding and hot-pressing. This might sound trivial, but such matters are crucial if nanotechnology is to be applied in

MICROFLUIDICS

Clicks and chips

The words 'organic chemistry' tend to conjure up images of large bubbling flasks, brightly coloured test tubes and explosive reagents. Although most chemists still use flasks for their reactions, a growing number of them make their molecules using miniaturized devices. These 'labs on chips' require only tiny quantities of reagents, thus reducing cost, producing less waste and cutting down the time needed to perform a reaction and to analyse its products.

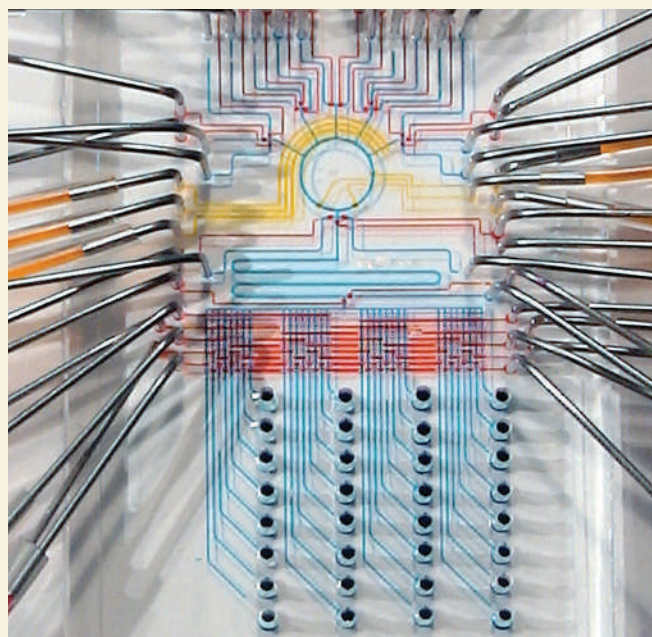
Although it is quite easy to perform a single reaction on a chip, it is much harder to carry out multiple reactions in the same device, making several different molecules. But now Hartmuth C. Kolb, Hsian-Rong Tseng and their colleagues have developed a device (pictured) that can perform 32 reactions at the same time. The group reports its results in *Angewandte Chemie* (doi:10.1002/anie.200601677).

The authors miniaturized a

technique known as 'in situ click chemistry', which can be used to identify high-affinity inhibitors for enzymes. These inhibitors can then be used in other biological experiments: for example, the molecules can be used to block the enzyme's 'active site' to see how it works, or to elucidate its cellular role.

Two compounds — one containing an azide group, and another containing an acetylene group — are combined in a reaction vessel with the target enzyme. Molecules of both compounds may enter the enzyme's active site, and orient themselves so that the azide group and the acetylene group fit comfortably inside. If the two groups align favourably, then 'click', they react to form a five-membered ring. Because this click reaction occurs in the active site of the enzyme, the product usually binds very tightly to that enzyme.

By miniaturizing this process on a chip, the authors were able to run 32



WILEY-VCH

reactions at the same time to find inhibitors for a well-known target enzyme (bovine carbonic anhydrase II) using much less of the protein than would have been needed in a traditional, microtitre-based procedure.

Many interesting proteins are

notoriously difficult to obtain in large quantities, preventing their use in biochemical assays for inhibitors. This chip enables such proteins to be screened at last, and may open up many areas for biological and medicinal study.

Joshua M. Finkelstein

the real world. Previous work using a similar process of surface modification to produce thin films of graphite did not have this advantage⁴. Thin films, and other kinds of graphite sheet, could be used for specialized applications, in electronic devices or for certain high-performance nanocomposites⁵.

Stankovich and colleagues' efforts³ resulted in composites with intriguing properties. Only 0.1% by volume of graphene in the composite is required for electrical conduction — this is the lowest threshold observed for conductive polymer composite materials, except for those using carbon nanotubes. The conductivity rapidly increases by incorporating more graphene, reaching 1 siemen per metre at a loading of 2.5% by volume; conductivities in the range of 0.1 S m^{-1} are sufficient for many applications.

The graphene composites could be very useful: for example, in the manufacture of fuselages for aircraft, which must combine low weight, high strength and electrical conductivity. This last property is necessary for protection against lightning strikes while in flight. Nevertheless, the conductivities of these composites³ are still several orders of magnitude lower than those of the best examples of nanotube mats (which are made entirely of nanotubes). They are also lower than the conductivity of graphite itself, or that of other nanotube composites^{6–8}. The positive trade-offs for graphene-sheet composites are low cost and the plentiful supply of graphite.

The small amount of graphene required in the composite³ for conductivity results from a high probability of sheet-to-sheet contact even at relatively low graphene loading, and from the conductive highway formed by the overlapping electron clouds of adjacent carbon atoms. However, high conductivities, rivaling those of individual nanotubes or thin nanotube

films, will not be achieved with these composites. This is because of the limited degree to which the phenyl-isocyanate-modified sheets mix in the polymer solution; the large number of defects in the carbon layers; and the slow rate of electron tunnelling through gaps between the sheets.

The technology described by Stankovich and colleagues has two main advantages. The first is its ease of use for large-scale industrial applications where the conductivity of carbon fibres is insufficient, but where carbon nanotubes would be too expensive. The second is its applicability to a variety of polymers. The modification of graphite oxide by phenyl isocyanate should be considered as a proof-of-concept demonstration. The phenyl group could be replaced by other groups compatible with different polymers. This opens up a wide area of research that could lead to a large family of composites with differing properties. Clearly, the next step is to determine the mechanical properties of the graphene composites, and to see whether they can compete with nanotube-based materials. ■

Nicolas A. Kotov is in the Department of Chemical Engineering, University of Michigan, 2300 Hayward Street, Ann Arbor, Michigan 48109-2136, USA.

e-mail: kotov@umich.edu

1. www.sciencenews.org/articles/20021005/bob9.asp
2. Coleman, J. N., Khan, U. & Gun'ko, Y. K. *Adv. Mater.* **18**, 689–706 (2006).
3. Stankovich, S. et al. *Nature* **442**, 282–286 (2006).
4. Kotov, N. A., Dekany, J. H. & Fendler, J. H. *Adv. Mater.* **8**, 637–641 (1996).
5. Mamedov, A. A. et al. *Nature Mater.* **1**, 190–194 (2002).
6. Zhang, M., Atkinson, K. R. & Baughman, R. H. *Science* **306**, 1358–1361 (2004).
7. Shaffer, M. S. P. & Windle, A. H. *Adv. Mater.* **11**, 937–941 (1999).
8. Yang, Y., Gupta, M. C., Dudley, K. L. & Lawrence, R. W. *Nanotechnology* **15**, 1545–1548 (2004).

STRUCTURAL BIOLOGY

Proteins downhill all the way

Jeffery W. Kelly

The hundreds of hydrogen atoms in a protein can be used as reporters to describe how the protein folds into and out of shape. The results challenge the dogma that this is always an all-or-nothing process.

The three-dimensional structures of proteins govern their activity, yet we know far less than we would like to about how these molecules fold into shape. Proteins use an intricate network of weak, non-covalent interactions to acquire the folded state¹. Conventional wisdom states that protein folding is a highly cooperative process — proteins are either completely folded or completely unfolded. This all-or-nothing model is convenient because it enables spectroscopic data to be

converted into thermodynamic data, simply by measuring the distribution of folded and unfolded molecules at equilibrium. But is this model always correct? Muñoz and colleagues (page 317 of this issue)² use nuclear magnetic resonance (NMR) spectroscopy to follow the unfolding of an all-helical protein known as BBL.

BBL folds in a 'downhill' fashion — that is, the process is characterized by very low energy barriers between the folded and unfolded



50 YEARS AGO

International conferences can be very stimulating affairs for those who attend, and the discussions, in particular, open up entirely new lines of thought. Publication of the proceedings can extend the stimulus to a much wider circle of workers, but only if the publication follows close on the heels of the conference itself... The proceedings of the symposium on nutritive aspects of preserved food... have been published approximately eighteen months after the conference took place; despite this delay, many of the papers are still badly in need of editing, the English sometimes being so poor that a sentence must be read several times over before its meaning can be grasped. From *Nature* 21 July 1956.

100 YEARS AGO

"The day of the week for any date" — We assign a number for each month in accordance with the old style, beginning with March, so that the last four months are numbered according to their Latin names, as follows: January, 0; February or March, 1; April, 2; May, 3; June, 4; July, 5; August, 6; September, 7; October, 8; November, 9; December, 10; next January 11; next February, 12. For a Leap Year, January and February must count as 11 and 12 respectively in the preceding year. It is only in dealing with the month-number that anything not straightforward and obvious is involved. The rule then runs as follows:

- A. For the century: divide by 4 and calculate 5 times the remainder.
- B. For the year: add to the number the quotient obtained from divisor 4.
- C. For the month: multiply by 4, and negate the units digit (i.e. subtract instead of adding it).
- D. For the day retain the number unchanged.

Then add together the results A, B, C, D (casting out sevens, of course, as you proceed), and the result gives the required day of the week...

Examples—1815, June 18 (Battle of Waterloo).

- A. For century: $2 \times 5 = 10 \equiv 3$
- B. For year: $15 + 3 = 18 \equiv 4$
- C. For month: 4×4 gives $10 - 6 = 4 \equiv 4$
- D. For day: 18 $\equiv 4$

$A + B + C + D = 15 \equiv 1$, i.e. Sunday
From *Nature* 19 July 1906.

50 & 100 YEARS AGO

states^{3–8}. When downhill folders are subjected to conditions that predispose them to unfold, such as heat stress, they are predicted to unfold gradually^{5,6}. The authors' findings² support this prediction, and suggest that protein folding is not necessarily an all-or-nothing process. Because unfolding is perceived to be the reverse of folding, the former process provides insight into the latter.

Every hydrogen atom in a protein is in a unique electronic environment because of the chemical groups surrounding it. As a result, each hydrogen atom gives rise to a distinct signal in an NMR spectrum, and the frequency of this signal changes when the protein unfolds, as the environment around the hydrogen atom alters. When heated to induce unfolding, BBL adopts many conformations that interchange rapidly, so the NMR frequency of each hydrogen atom reports on the extent of local unfolding.

Muñoz and colleagues² plotted the NMR frequency changes of 158 hydrogen atoms in BBL as a function of temperature, producing 158 atomic-resolution graphs, known as unfolding curves. These atomic unfolding curves were not superimposable. Instead, the midpoint temperatures of the curves spanned a temperature range of 60 K. This demonstrates that, at any given temperature, some hydrogen atoms were in a folded environment whereas others were in an unfolded environment. This observation cannot be explained by the existence of only folded or unfolded BBL molecules².

Low-resolution unfolding curves have typically been obtained by methods such as circular dichroism (CD), which detects changes in the amount of polarized ultraviolet light absorbed by a protein as it folds. Interestingly, the average of the 158 NMR-based, atomic-resolution unfolding curves² is very similar to the CD-derived BBL unfolding curve. In other words, the great complexity of unfolding observed at the atomic level is reduced to a simple, two-state-like unfolding curve when monitored at low resolution. At the global unfolding mid-point there is an almost uniform distribution of atoms with environments ranging from fully folded to fully unfolded. This belies the notion that all proteins can exist only in completely folded or completely unfolded forms. Again, this is a surprising conclusion, as the all-or-nothing assumption is integral to the equations used to extract thermodynamic data from unfolding curves.

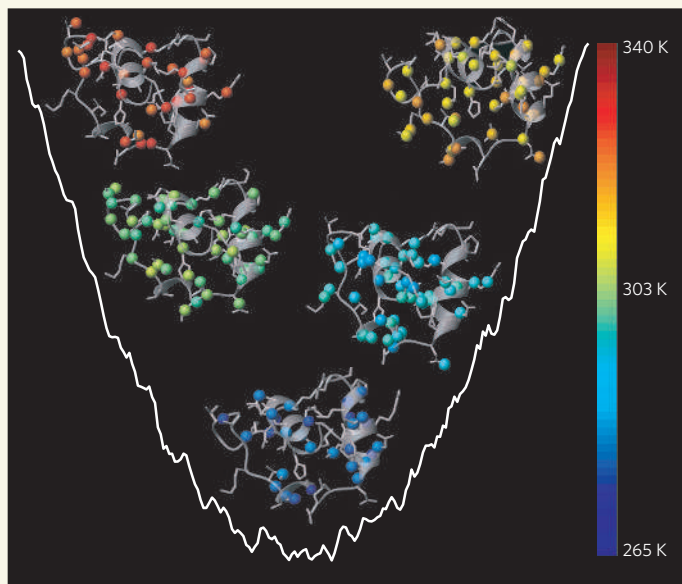


Figure 1 | An energy landscape for the folding of the all-helical protein BBL. This is a two-dimensional slice of a three-dimensional energy surface, with the conformations formed at higher temperatures being higher-energy structures. BBL is referred to as a downhill folder because of the very low energy barriers that exist between conformational states. The landscape theory of protein folding implicates numerous paths and thus mechanisms for passing from the unfolded to the folded state. There are many possible unfolded states, and therefore many starting points for folding (of which we are seeing just two), but only one folded state at the bottom of the well. Muñoz and colleagues² use NMR to provide an unprecedented view of the unfolding of BBL. They reveal that downhill-folding proteins acquire structure gradually, not in an all-or-nothing or highly cooperative process; thus, families of partly cooperative mechanisms must be considered statistically to explain downhill folding.

So BBL unfolds in a gradual fashion (Fig. 1). This means that the midpoint of the low-resolution unfolding curve for BBL corresponds to a state of maximal conformational heterogeneity, as seen in the NMR studies on individual hydrogen atoms². Because of statistical considerations, many of the atomic unfolding curves are bound to be similar to the low-resolution unfolding curve, but this does not prove the all-or-nothing theory of protein folding for BBL.

Muñoz and colleagues' work provides fresh evidence that protein folding actually proceeds through a complex mixture of numerous pathways, as predicted by the landscape view of protein folding³ (Fig. 1). Just as a skier can take an almost infinite number of paths to get from the top of the slope to the chair-lift, by analogy, the landscape theory of protein folding implicates numerous paths and thus mechanisms for passing from the unfolded to the folded state.

Groups of hydrogen atoms were observed in BBL that experience unfolding at similar temperatures². Interestingly, these groups of atoms are not necessarily close to each other in the protein, which implies that there is no defined sequence of local unfolding events. In fact, the authors' data suggest that, upon heating, a large fraction of the hydrogen bonds in BBL break first. This is followed by the loss of interactions between the side-chains of the protein's amino acids — a significant event, as these interactions define the overall shape of the

folded protein. At higher temperatures, alterations to the backbone helical conformations are observed. Even so, backbone conformational preferences persist at high temperatures, demonstrating the lack of strong coupling between side-chain interactions and the overall shape of the backbone.

Muñoz and colleagues² quantify the extent of all-or-nothing folding in BBL by using a parameter they call the mean thermodynamic coupling index (MTCI), which is based on the similarity between the atomic-resolution unfolding transitions. The MTCI approaches infinity for an all-or-nothing folding process, and approaches zero for a set of totally uncoupled atomic transitions. The MTCI of BBL is around $1.2RT$, where R is the ideal-gas constant and T is temperature. This is scarcely an all-or-nothing transition — in fact, it is remarkable that such a poorly cooperative process is sufficient to cause folding.

The authors show that hundreds of atomic probes within BBL provide an unprecedented view of downhill protein fold-

ing. The conclusions of this study may apply only to proteins with energy barriers to folding smaller than around $3RT$. Nevertheless, this statistical approach towards understanding protein folding will probably be applied to other small, rapidly folding proteins, previously assumed to be two-state folders, to discern the generality of this finding⁹. This paper is significant not only for the above reasons, but also because it puts the thermodynamic approach to downhill folding on a very strong footing. And it reminds us that we should not be so quick to ignore experimental data that do not fit the classic model of all-or-nothing folding. ■

Jeffery W. Kelly is in the Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA.
e-mail: jkelly@scripps.edu

1. Fersht, A. R., Matouschek, A. & Serrano, L. *J. Mol. Biol.* **224**, 771–782 (1992).
2. Sadqi, M., Fushman, D. & Muñoz, V. *Nature* **442**, 317–321 (2006).
3. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. *Proteins* **21**, 167–195 (1995).
4. Sabelko, J., Ervin, J. & Gruebele, M. *Proc. Natl Acad. Sci. USA* **96**, 6031–6036 (1999).
5. Muñoz, V. *Int. J. Quantum Chem.* **90**, 1522–1528 (2002).
6. Garcia-Mira, M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M. & Muñoz, V. *Science* **298**, 2191–2195 (2002).
7. Yang, W. Y. & Gruebele, M. *Nature* **423**, 193–197 (2003).
8. Kubelka, J., Hofrichter, J. & Eaton, W. A. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).
9. Jager, M. *et al. Proc. Natl Acad. Sci. USA* **103**, 10648–10653 (2006).

BRIEF COMMUNICATIONS

Packing grains by thermal cycling

Small temperature changes can affect the packing of granular materials without mechanical disturbance.

A long-standing problem in managing the behaviour of a collection of solid grains concerns the nature of the grain packing¹, a property that is typically controlled by how the grains are poured or shaken. Here we show that a systematic and controllable increase in granular packing can be induced by simply raising and then lowering the temperature, without the input of mechanical energy. This thermal processing may have important practical implications for the handling and storage of granular materials.

The packing fraction of a granular material is defined as the fraction of sample volume that is filled by grains rather than by empty space, and typically varies between 57% and 64% for randomly arranged, spherical grains and even more widely for other grain shapes. The density of packing can be increased by vibration or tapping, which induce small rearrangements that allow the grains to settle^{2–4}.

When a granular material is heated, the grains and their container both undergo thermal expansion. This can lead to settling because of the metastable nature of disordered grain configurations (especially if the grains and their container are made of different materials), and such settling should not be reversible upon cooling to ambient temperature. It has been shown that temperature changes affect silos in industrial settings^{5,6} and the stress state of a granular pile^{7,8}. But the grain dynamics induced by thermal cycling have not been analysed until now.

We examined the change in packing fraction for glass spheres contained in vertical plastic cylinders in response to thermal cycling (using both single thermal cycles from room temperature and repeated cycles over the same temperature range; for methods, see supplementary information). We found that there was a clear increase in packing even for a single cycle to 10 °C above ambient room temperature (Fig. 1a). The results were not affected by the height to which the cylinders were filled (to within $\pm 10\%$), the heating rate, or the time spent at the cycle temperature after thermal equilibrium was reached. They changed only slightly ($< 20\%$) if the cylinder diameter was changed by an order of magnitude (Fig. 1b). This latter result is physically sensible, because the expansion of the grains and of the container scales with the size of the sample.

The packing fraction continues to increase over multiple thermal cycles (Fig. 1c). The

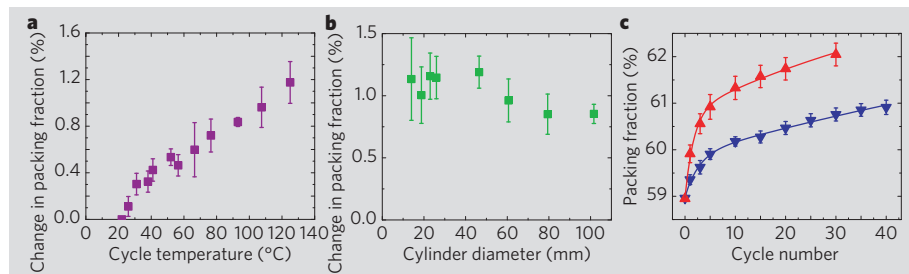


Figure 1 | Change in packing fraction with thermal cycling for glass spheres in a plastic cylinder. **a, b,** Change in packing fraction as **a**, a function of cycle temperature from room temperature for a single cycle (cylinder diameter, 60.6 mm) and **b**, a function of cylinder diameter for a single cycle (cycle temperature, 107 ± 2 °C). **c**, Change in packing fraction after multiple thermal cycles (cycle temperatures: red, 107 ± 2 °C; blue, 41 ± 1 °C). Lines represent fits of the data to a double-relaxation model (ref. 9; for details, see supplementary information). Error bars represent the standard deviation determined from several repeated measurements (**a, b**, 12 or more measurements; **c**, 6 measurements).

increase can be described by a double-exponential density-relaxation model, consistent with a combination of large-scale (relaxation of granular blocks) and small-scale (relaxation of individual particles) rearrangements (for details of the model and fits to the data, see supplementary information)⁹. Although the limited range of the experimental data restricts thorough testing of this double-relaxation model, the ‘time constants’ of the two different relaxation times observed for each cycle temperature do increase with decreasing cycle temperature — consistent with smaller thermal expansion, as more cycles are needed to achieve the same change in packing fraction at a lower cycle temperature.

The primary cause of the changes in packing fraction noted here is probably the difference between the thermal expansion of the container and of the grains. This explanation is confirmed by our observation (results not shown) of similar changes in packing of plastic spheres in glass cylinders (where the grains expand more than the container), and of smaller changes in packing for glass spheres contained in glass cylinders.

Our results indicate that there may be many thermal effects in granular media, analogous to the effects of vibration. For example, a geophysical form of granular segregation (‘stone heave’) has been associated with thermal effects¹⁰. Temperature changes can also cause granular pressure to increase in outdoor silos with each diurnal cycle — potentially leading to catastrophic failure of the silo⁶. Our results demonstrate that thermal cycling can provide an almost ‘adiabatic’ alternative to mechanical agitation for altering grain packing.

K. Chen, J. Cole, C. Conger, J. Draskovic, M. Lohr, K. Klein, T. Scheidemantel, P. Schiffer
Department of Physics and Materials Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA
e-mail: schiffer@phys.psu.edu

1. Donev, A. *et al.* *Science* **303**, 990–993 (2004).
2. Nowak, E. R., Knight, J. B., Ben-Naim, E., Jaeger, H. M. & Nagel, S. R. *Phys. Rev. E* **57**, 1971–1982 (1998).
3. Philippe, P. & Bideau, D. *Europhys. Lett.* **60**, 677–683 (2002).
4. Richard, P., Nicodemi, M., Delannay, R., Ribière, P. & Bideau, D. *Nature Mater.* **4**, 121–128 (2005).
5. Blight, G. E. *Int. J. Bulk Solids Storage* **1**, 1–7 (1985).
6. *Flow of Solids 1–2* (Fall, 2001); www.jenike.com/pages/education/cases/pdf_docs/O2jcase.pdf
7. Liu, C. H. *Phys. Rev. B* **50**, 782–794 (1994).
8. Puri, V. M., Zhang, Q. & Manbeck, H. B. *Int. J. Bulk Solids Storage* **2**, 1–7 (1986).
9. Barker, G. C. & Mehta, A. *Phys. Rev. E* **47**, 184–188 (1993).
10. Viklander, P. *Cold Region Sci. Technol.* **27**, 141–152 (1997).

Supplementary information accompanies this communication on Nature’s website.

Received 18 January; accepted 15 June 2006.

Competing financial interests: declared none.
doi:10.1038/442257a

CORRIGENDUM

Animal communication: Complex call production in the túngara frog

M. Gridi-Papp, A. S. Rand, M. J. Ryan
Nature **441**, 38 (2006)

The bar coloration as specified in Fig. 1a of this communication is misleading: it should indicate the relative amplitude of odd-to-even harmonics before (red bars) and after (blue bars) treatment for the ‘chuck’ and ‘whine’ call components.

doi:10.1038/442257b

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

EARTH SCIENCE

Is earthquake rupture deterministic?

Arising from: E. L. Olson & R. M. Allen *Nature* 438, 212–215 (2005)

It is essential in an earthquake early-warning system to be able rapidly to determine the size and location of an earthquake. Olson and Allen¹ claim that the size of large earthquakes (magnitude (M) greater than 5.5) can be estimated from the seismic energy radiated during the first several seconds of fault rupture, implying that the earthquake process is deterministic. But here we analyse waveform data from more than 50 events ($M \geq 6.0$) recorded by the Japanese Hi-net seismic network and find no evidence that earthquake magnitude can be estimated before the rupture has completed. This bears on the difficult problem of understanding the physics of the earthquake process².

Investigations^{1,3} of seismic energy radiated during the initial fault rupture have suggested that the earthquake process may be deterministic, as these data apparently provide an estimate of the size of the event. The method of analysis was to find a velocity spectral peak in the first several seconds of P-wave seismic data. A peak indicates that there is a predominant period, $\tau_p^{\max}$, in the seismogram, which was found to scale with earthquake magnitude, suggesting that the magnitude of larger earthquakes can be estimated before rupture

has completed^{1,3}. For events smaller than about $M = 6.0$, several seconds of P-wave data contain almost the entire history of rupture, and conventional models of the earthquake process^{4,5} predict that the spectral peak will scale with magnitude, as is observed^{3,6}.

For events of magnitude $M \geq 6.0$, however, the claim by Olson and Allen that a spectral peak from a short — compared with rupture duration — time series predicts the earthquake's magnitude has necessarily resulted in speculation about the initiation process and the physical state of the fault. The authors suggest¹ that an earthquake develops into a large event if fault slip at the onset of rupture provides enough energy for continued propagation across fault heterogeneities, which could inhibit rupture and thereby result in a smaller event. This large slip would cause a velocity peak in the initial P-waveform data and would scale with earthquake size. It is not clear, however, what mechanism controls the information transfer between the heterogeneities across tens to hundreds of kilometres of fault zone and the initial slip. Also, waveform data have been used to model the fault slip⁷ and the largest amount of slip does not necessarily coincide with the location of the initial rupture.

A high-quality seismic network (Hi-net), comprising some 700 stations that have seismometers set in boreholes at depths of more than 100 metres, was installed in Japan after the 1995 Kobe earthquake. In an earthquake-prone country, the ability to predict an earthquake's size ($M > 6.0$) from initial P-wave data would be invaluable in an earthquake-warning system⁸, as not only could the hypocentre be located from the P-wave arrival times^{8,9}, but the (eventual) size could also be quickly estimated.

We analysed Hi-net data for 52 Japanese earthquakes with a magnitude range of $6.0 \leq M \leq 8.0$, as determined by the Japan Meteorological Agency¹⁰. For each event, and using the same method of analysis as Olson and Allen¹, we calculated the predominant period, τ_p , in the velocity spectrum of the P-wave arrival data from the five stations closest to the earthquake and then found the maximum value, $\tau_p^{\max}$; these values were then averaged (Fig. 1). No trend is evident in Fig. 1 as the correlation coefficient is not statistically significant and the probability of a chance occurrence is high. Our results show that the eventual size of Japanese earthquakes in this magnitude range cannot be determined from

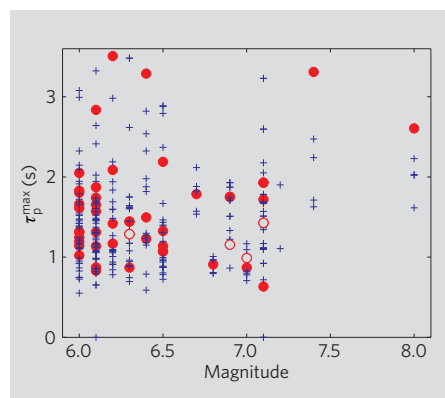


Figure 1 | Plot of $\tau_p^{\max}$, the peak value of the predominant period in P-wave arrival data, versus magnitude for large ($M > 6.0$) earthquakes recorded by the Hi-net seismic array in Japan. Filled circles show the mean values of $\tau_p^{\max}$ that were estimated from the five stations (cross symbols) that were closest to the epicentre of each earthquake. Regression analysis gives a correlation coefficient ($r = 0.16$) that is not statistically significant and has a high probability ($P = 0.26$) of occurring by chance from a random distribution. Open circles are values from the four earthquakes used in Fig. 2. The Hi-net array has an inter-station spacing of 20–25 km and employs three-component seismometers; however, only the vertical component is used in this instance.

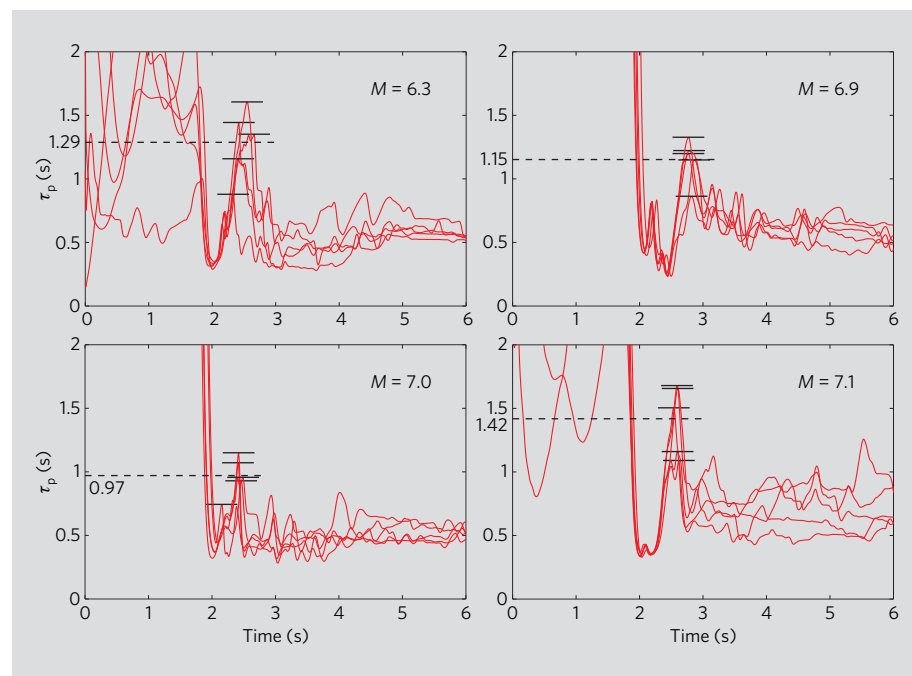


Figure 2 | Time history of τ_p analysis, the predominant period of P-wave arrival data, for four earthquakes in the magnitude (M) range of 6.3 to 7.1. The five traces for each event are from the five stations that were nearest to the epicentre. A solid horizontal bar is centred over the position of the $\tau_p^{\max}$ peak for each station and the dashed line is the average of these peak values (open circles in Fig. 1). The P-wave arrival here is always at 2 s and the background noise in the seismogram may produce large values of τ_p before the arrival of the seismic energy.

just several seconds of P-wave data, a position that disagrees with Olson and Allen¹.

Closer inspection of Olson and Allen's plot of $\tau_p^{\max}$ versus magnitude (Fig. 3 in ref. 1) reveals a distinct break in the linear trend of the data at about $M = 5.5$. Below this magnitude, a scaling is observed but, as already noted, such scaling is expected from conventional seismic theory. Above $M = 5.5$, scaling is less evident¹, but neither the correlation coefficient nor the statistical significance was calculated by the authors for these larger events. We calculate the correlation $r = 0.34$ for this section of the trend, with a probability of 0.074 of chance occurrence; these values therefore represent a case of borderline statistical significance and should be viewed with caution, particularly given the important implications. The larger sample used here (about 2.3 times the size of that of Olson and Allen¹) reveals no

trend and hence we cannot support the claim that the first several seconds of rupture energy can be used to estimate the eventual size of a large earthquake.

The individual station estimates of τ_p for four larger earthquakes (Fig. 2) are also evidence for the poor predictability of $\tau_p^{\max}$ analysis. The results from station to station are generally consistent for $\tau_p^{\max}$, but no systematic increase in $\tau_p^{\max}$ with earthquake size, nor the time at which it occurs, is found for any of these events.

It would be particularly advantageous if the magnitude of a large ($M > 6.0$) earthquake could be estimated from just several seconds of P-wave data, because a real-time warning could then be issued before the later arrival of the highly destructive S waves. However, our $\tau_p^{\max}$ analysis of Hi-net data does not seem to provide this information in advance.

Paul Rydelek*, Shigeki Horiuchi†

*Center for Earthquake Research and Information, The University of Memphis, Memphis, Tennessee 38152, USA

e-mail: prydelek@memphis.edu

†National Institute for Earth Science and Disaster Prevention, Tsukuba 305-0006, Japan

1. Olson, E. L. & Allen, R. M. *Nature* **438**, 212–215 (2005).
2. Sato, K. & Mori, J. *J. Geophys. Res.* **111**, B05307 (2006).
3. Allen, R. M. & Kanamori, H. *Science* **300**, 786–789 (2003).
4. Kanamori, H. *J. Geophys. Res.* **82**, 2981–2987 (1977).
5. Sacks, I. S. & Rydelek, P. A. *Bull. Seismol. Soc. Am.* **85**, 808–813 (1995).
6. Kamigaichi, O. *Jpn Assoc. Earthquake Eng.* **4**, 134–137 (2004).
7. Wald, D. J., Heaton, T. H. & Hudnut, K. W. *Bull. Seismol. Soc. Am.* **86**, S49–S70 (1996).
8. Horiuchi, S., Negishi, H., Abe, K., Kamimura, A. & Fujinawa, Y. *Bull. Seismol. Soc. Am.* **95**, 708–718 (2005).
9. Rydelek, P. A. & Pujol, J. *Bull. Seismol. Soc. Am.* **94**, 1546–1550 (2004).
10. Tsuboi, C. *Zisin* **7**, 185–193 (1954).

doi:10.1038/nature04963

EARTH SCIENCE

Olson & Allen reply

Replying to: P. Rydelek & S. Horiuchi *Nature* **442**, doi:10.1038/nature04963 (2006)

Rydelek and Horiuchi¹ present an analysis of earthquakes in Japan whose magnitudes (M) ranged from 6.0 to 8.0. They determine $\tau_p^{\max}$, the peak predominant period for P-wave arrival, for the five nearest Hi-net seismic-network stations and plot the event-averaged $\tau_p^{\max}$ versus magnitude. They find no statistically significant trend in their data.

The scaling relation between $\tau_p^{\max}$ and magnitude that we presented² was derived for earthquakes with magnitudes ranging from 3.0 to 8.3. In the analysis of Rydelek and Horiuchi¹, all but two of the events are in the magnitude range of 6.0 to 7.1. When a large magnitude range is used, the scaling relation is clear, as shown in Fig. 3a of ref. 2. However, there is also scatter in the magnitude of events with the same $\tau_p^{\max}$ observation. The full range of variability in the event-averaged $\tau_p^{\max}$ that we observed is ± 1 magnitude unit. Given the narrow magnitude range used by Rydelek and Horiuchi¹, and the variability in $\tau_p^{\max}$ observations as noted by us, it is not possible — and, indeed, we do not expect — to observe a significant trend. Rydelek and Horiuchi¹ cannot therefore call into question the existence of a scaling relation between $\tau_p^{\max}$ and magnitude based on their data set. The uncertainty in their event-averaged $\tau_p^{\max}$ observations¹ is also increased by their use of only five stations per event; we used all available data within 100 km of each event, which is an average of 26 stations per event.

Our interpretation that earthquake rupture is deterministic is based on the observation that the magnitude can be estimated before rupture is complete. Rydelek and Horiuchi¹ have explained that their rationale for using

only $M \geq 6.0$ events is that the rupture duration for events of $M = 6.0$ is about 4 s, claiming that our τ_p analysis uses 4 s of data, and therefore that only events with $M \geq 6.0$ can be used to determine whether rupture is deterministic (P. Rydelek and S. Horiuchi, personal communication).

Our τ_p analysis does not use 4 s of data: it uses a maximum of 4 s of data. The actual amount of data used is shown in Fig. 3b of ref. 2, which plots the delay time, τ_d , at which $\tau_p^{\max}$ is observed with respect to the P-wave trigger, together with the typical rupture duration. For all events that plot below the rupture-duration line, the magnitude can be estimated before rupture cessation, suggesting a deterministic process. This is the case for 58 of the 72 earthquakes in our study. These 'deterministic' events have a magnitude range of $3.3 \leq M \leq 8.3$ and include all events with $M \geq 4.0$.

Rydelek and Horiuchi¹ also point out a "break" in the linear trend in $\tau_p^{\max}$ at $M = 5.5$. This is a visual artefact resulting from the criteria used to select the data set. The apparent break results from the fact that all available earthquakes from southern California were combined with $M \geq 6.0$ events from other regions around the world. This resulted in fewer events in the $5.5 \leq M \leq 6.0$ range. It should also be noted that the uncertainty in a magnitude estimate derived from a $\tau_p^{\max}$ scaling relation for a specific region, California for example, is lower than when combining observations from a global data set.

Our interpretation for the observed scaling relation is that a stronger initiation will lead to a rupture that is statistically more likely to propagate further across a heterogeneous fault

plane. This has been suggested previously^{3–5} and does not require the largest amount of slip to be at the hypocentre. Although the hypothesis is controversial^{6,7}, another study⁸ shows that the region of largest slip is statistically near the hypocentre, implying that the location of rupture initiation and peak slip are not independent of one another.

Beyond its implications for the physics of earthquake rupture, rapid magnitude determination can also be used in earthquake-warning systems to provide seconds to tens of seconds of warning before the onset of severe ground-shaking^{9,10}. The Japan Meteorological Agency currently operates an early-warning system that estimates the location and magnitude of an earthquake using P-wave arrivals¹¹. Starting in June 2006, any organization can apply to receive early warnings of earthquakes.

Erik L. Olson*, Richard M. Allen†

*Department of Geology and Geophysics, University of Wisconsin, Madison, Wisconsin 53706, USA

†Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA
e-mail: rallen@berkeley.edu

1. Rydelek, P. & Horiuchi, S. *Nature* **442**, doi:10.1038/nature04963 (2006).
2. Olson, E. L. & Allen, R. M. *Nature* **438**, 212–215 (2005).
3. Ellsworth, W. L. & Beroza, G. C. *Science* **268**, 851–855 (1995).
4. Beroza, G. C. & Ellsworth, W. L. *Tectonophysics* **261**, 209–227 (1996).
5. Heaton, T. H. *Phys. Earth Planet. Int.* **64**, 1–20 (1990).
6. Kilb, D. & Gomberg, J. *J. Seismol.* **3**, 409–420 (1999).
7. Mori, J. & Kanamori, H. *Geophys. Res. Lett.* **23**, 2437–2440 (1996).
8. Mai, P. M., Spudich, P. & Boatwright, J. *Bull. Seismol. Soc. Am.* **95**, 965–980 (2005).
9. Allen, R. M. *Seismol. Res. Lett.* **77**, 371–376 (2006).
10. Lockman, A. B. & Allen, R. M. *Bull. Seismol. Soc. Am.* **95**, 2029–2039 (2005).
11. Kamigaichi, O. *Jpn Assoc. Earthquake Eng.* **4**, no.3 (2004).

doi:10.1038/nature04964

Ecological networks and their fragility

José M. Montoya^{1,2}, Stuart L. Pimm³ & Ricard V. Solé^{2,4}

Darwin used the metaphor of a 'tangled bank' to describe the complex interactions between species. Those interactions are varied: they can be antagonistic ones involving predation, herbivory and parasitism, or mutualistic ones, such as those involving the pollination of flowers by insects. Moreover, the metaphor hints that the interactions may be complex to the point of being impossible to understand. All interactions can be visualized as ecological networks, in which species are linked together, either directly or indirectly through intermediate species. Ecological networks, although complex, have well defined patterns that both illuminate the ecological mechanisms underlying them and promise a better understanding of the relationship between complexity and ecological stability.

A food web maps which species eat which other species^{1–3}. Such trophic interactions—henceforth 'links'—are typically antagonistic predator–prey ones, but we expand the discussion to include webs of mutualists^{4,5}, including, for example, the interactions of flowers and their pollinators, and fruits and their seed dispersers.

Ecological networks are complex, with each species typically closely linked to all others, either directly or indirectly^{6,7}. For example, links tend to be 'nested'^{8,9}—that is, the diet of the most specialized species is a subset of the diet of the next more generalized species, and its diet a subset of the next more generalized, and so on. The most generalized species may include most of the prey species present in its diet. Within this pattern of close links to other species, there are reasons to expect even denser clusters of links. For example, coevolution is expected to generate clustered links that stem from the reciprocal mutual specialization between plants and their pollinators or seed dispersers^{4,5}. The general existence and underlying causes of dense clusters is, however, still a matter of debate^{3,10}.

In this Review, we compare ecological networks to non-ecological networks, and consider their similarities, differences and underlying causes. Along with networks of interacting computers, genes or humans, ecological networks display well-defined, similar patterns of organization^{11–13}. On close inspection, ecological networks are unlike other networks. Their assembly follows different rules, and the processes of predation, competition and mutualism constrain them in unique ways. Other networks nonetheless help us understand why ecological ones are special in the constraints that apply to them and how they develop.

Finally, there is a paradox. Theory predicts that complex ecological networks are likely to be 'fragile' in various ways. For example, complexity affects the chance that species can coexist at a stable equilibrium. It also affects resilience (how fast populations recover from disturbances), the variability of population densities over time and the persistence of community composition. It should also affect the resistance to change, when species invade or are driven to extinction^{14–16}. Every species is closely linked to every other^{6,7}, so—metaphorically—when a tree falls in a rainforest, every species in that species-rich, complex system would seem to 'hear' that event quickly. Every disturbance buffets every other species, so how do species persist in this 'noisy' world? And which species will persist in the now extensively modified and increasingly species-poor world we are

creating for them? We address these questions in the penultimate section, where we highlight both the extent of our present understanding, and the important gaps in our knowledge.

Complexity

No species is distant from the most connected species, nor any top predator from a species at the base of the web^{2,6} (Fig. 1; Box 1 defines 'distance' and other technical terms). In seven species-rich webs, 80% and 97% of species are, respectively, within two or three links from each other⁷. Both the cascade model², and the niche model^{17,18} it inspired, predict short distances between top predators and plants or detritivores at the base of a web¹⁹. The cascade model posits nested diets, with the top predator potentially exploiting all the other species, the next predator exploiting all but the top predator, and so on.

Given that species in a web are typically closely connected, are there even denser patterns of links within each web? The first approach to this question considers detailed patterns, and can be applied to webs with few or many species. (Many early studies involved webs with few species. Box 2 considers the limitations on what data known food webs provide.)

Food webs straddling different habitats generate compartments corresponding to habitat boundaries²⁰. But no agreement exists when looking for compartments within a habitat²⁰. Some visual inspections for compartments find them^{21,22}, while others do not²³. Using statistics, one study²⁴ finds compartments within the food web of an estuary. Another²⁰ finds roughly equal numbers of more and less compartmented webs than expected by chance. A third study identified compartments in three of five species-rich food webs²⁵.

Analysing clusters is more tractable than detecting compartments. In a random web with S species and an average of k links per species, clustering scales as k/S (refs 12, 13). Generally, webs with many species and few links per species are more clustered than are random counterparts^{6,10}. Embedded in the most highly clustered webs, however, are two classes of systems long known³ and theoretically expected³ to have tight clusters of links. In aquatic systems, many fish are 'life-history omnivores' that feed from several trophic levels at various life stages—on phytoplankton after hatching to other piscivorous fish as adults—including those that ate them when young²⁶ (see Fig. 1). Tight clusters are common in host–parasitoid systems, where a parasitoid may feed on a host, but also hyper-parasitize other parasitoids^{23,27}.

¹School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK. ²ICREA-Complex Systems Laboratory, IMIM-UPF, Dr Aiguader 80, 08003 Barcelona, Spain. ³Nicholas School of the Environment and Earth Sciences, Duke University, Durham, North Carolina 27708, USA. ⁴Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA.

In sum, important classes of webs involve densely linked groups of species. Life-history omnivores dominate the marine ecosystems that cover most of the planet (as well as many freshwater ecosystems), whereas plant–herbivore–parasitoid systems account for the great majority of Earth's species²⁸. From these perspectives, densely clustered webs are nearer the norm, not exceptions. Moreover, compartments are likely to correspond to habitat boundaries, if only because of the difficulties of species feeding successfully in two or more habitats. This begs the question, not addressed here, of what constitutes a habitat for anything less subtle than major categories such as 'ocean, lake, and land'²⁰.

These exceptions noted, there is no clear consensus about compartments and clusters. Whether the 'species' that the food web depicts are actually species or taxonomically aggregated groups or 'trophic species'—species that apparently share the same (or similar) predators and prey^{2,3}—confounds the results. Krause *et al.*²⁵ provide the most complete analysis, and show that the detection of compartments requires new algorithms. Increasing taxonomic resolution and details on the strength of interactions increased their chances of finding compartments. These complexities inevitably create controversy. Its resolution will come from applying their ideas to other quantified and taxonomically specific webs.

A second approach applies to only species-rich webs and considers $P(k)$, the probability of a species having k links (Fig. 2). Power law, 'scale-free' distributions, $P(k) \sim k^{-\gamma}$ (we use $\sim$ here to mean 'follows a function'), describe the links between human social contacts and internet connections^{11–13}. They suggest a 'popularity principle' whereby the 'rich get richer'. As these non-ecological networks grow, nodes are likely to get more links the more they have already¹², and dense clusters develop. In these systems, the constant γ is typically a value between 2 and 3 (refs 12, 13).

Alternatively, many factors constrain a predator's diet, so truncating the linkage distribution, preventing the 'rich from getting richer' and suggesting an exponential trimming of linkages: $P(k) \sim e^{-k/\xi}$. How large is the truncation measured by the parameter ξ ? One can fit the combined model $P(k) \sim k^{-\gamma}e^{-k/\xi}$ (refs 29, 30). Intuitively, the value ξ measures the number of links below which power laws describe the systems and above which the number of links per species drops off more steeply (Fig. 2).

For a set of 86 mutualistic webs³⁰ ranging from 17 to nearly 1,000

species, this truncation is >4 times the average number of links per species in two-thirds of the examples. In these examples, the power-law exponent γ is typically ~ 1 ; thus, the distributions are substantially less steep than are the non-ecological networks discussed above. For the set of species-rich food webs (29 to 238 species), ξ is smaller, roughly equal to the average number of links per species^{6,10,31}, consistent with the distributions arising from the niche model and a generalized version of the cascade model³².

In sum, the distributions of links in food webs do not match other networks quantitatively³³, and only match qualitatively for small ranges of numbers of links. Moreover, the 'rich get richer' mechanism is at odds with ecological principles^{5,10}. For example, as more species of frugivore feed upon a fruit species, then competition for that fruit will increase, and the less likely another frugivore would include it in its diet and the more likely it would feed on other fruits instead.

Species with many links to other species in the web—the 'richest' ones in the analogy—may get that way simply by being the most abundant. Within a given trophic level, abundant species are more likely to receive the attention of predators, pollinators or frugivores, than are rare ones. Abundant predators probably feed on many prey species³⁴.

Perversely, across trophic levels, the larger a species' body size, the more species on which it can feed, and thus the higher its trophic level⁹. Large body size and high trophic level mean lower abundance^{9,35}, so across trophic levels it is the trophically most generalized species that are the rarest.

Whether within or across trophic levels, these patterns of strongly nested diets—whether by abundance or body size (or both)—do not imply a process of continuing community assembly. That said, Sugihara³⁶ and colleagues³⁷ do envision a process whereby species sequentially partition resources as they invade an ecological community. Rarer, trophically specialized species enter the community later than do generalists.

The major gap in knowledge is surely the scarcity of models exploring dynamics and rules for coexistence of large sets of species with trophic links suitably nested across trophic levels by body size (and associated attributes) and within trophic levels by abundance. Recent descriptions of food webs detailing body-size, abundance and trophic linkages^{9,35,38} provide a good base to develop these models.

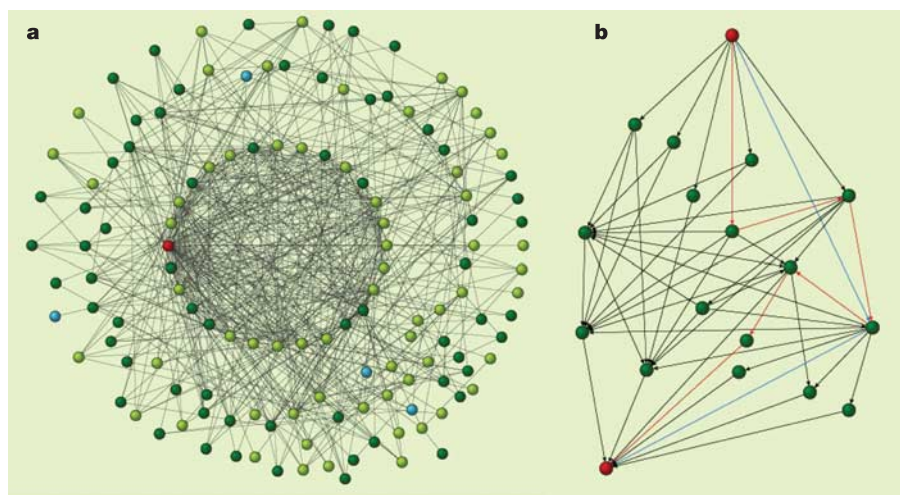


Figure 1 | Shortest paths in a complex food web. The Ythan estuary food web^{66,73}. **a**, Node colour indicates the length of the shortest path linking the most connected species (the flounder *Platichthys flesus*, in red) and each other species from the network. (The trophic direction of the links—what eats what—is ignored). Dark green, species are one link apart; light green, two links; and blue, three links. The central circle represents the densest sub-web⁷⁵, which consists of 28 species with at least 7 links with the rest of

the species from the sub-web. This sub-web contributes most to the observed clustering. **b**, Food chains between basal species of *Enteromorpha carbo* (red node at the bottom) and the top predator, the cormorant *Phalacrocorax carbo* (red node at the top). Links corresponding to the shortest path connecting them are in blue (2-links), and those corresponding to the longest food chain between these two species are in red (6-links).

Box 1 | Definitions

Complexity, often used informally, has a specific meaning—the average number of trophic links per species (sometimes called linkage density). **Connectance** is linkage density divided by the number of species in the web^{2,3}. **Linkage strength** is the magnitude of the effect of one species upon the growth rate of another⁵².

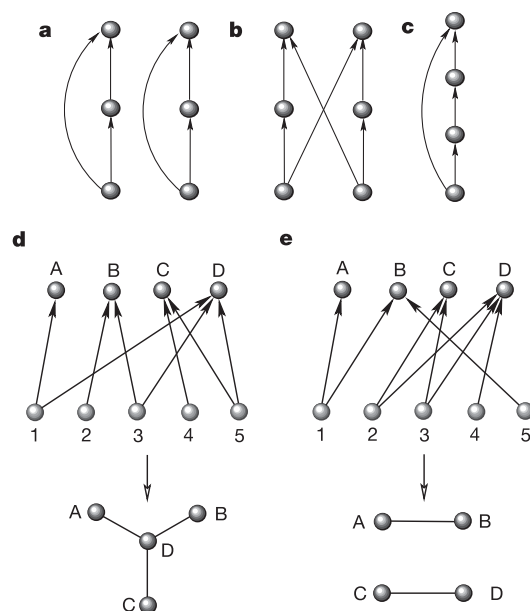
On theoretical grounds, May⁴⁶ conceived **blocks**, which Pimm and Lawton²⁰ defined operationally as **compartments**. Links connect species within a compartment, but not to species outside it (see Box Figure). In practice, there may be links outside the compartment, but these should be relatively few or relatively weak dynamically compared to links within it. We can also measure **clustering** as the fraction of species with a direct link to a focal species that are themselves directly linked^{65,66} (see Box Figure).

An **omnivore** feeds on more than one trophic level⁶⁷ and so ‘messes up’ classifying species into distinct levels³¹. With two important classes of exceptions, early studies found omnivory to be statistically unusual in many systems³—justifying the long-held notion of distinct trophic levels. Recent studies confirm this: half the species of a set of species-rich food webs can be unambiguously assigned to a discrete trophic level¹⁹. For a given complexity, webs with many within-food-chain omnivores tend to be compartmented²⁰ and clustered, though the relationship is not exact (see Box Figure).

A **guild** of predators feeds on the same prey species, distinct from those other guilds exploit⁶⁸. Co-evolution expects **reciprocal specialization** between pollinators and their flowers and between fruit and their dispersers (see Box Figure).

Network studies report the statistical distributions of the numbers of links per species^{11–13,61}, and ask if ecological networks are **small worlds**^{6,11,65}—where no species are ‘distant’ from each other.

Distance (also called **shortest path**) is the minimum number of links connecting two species: one, if A feeds on B; two, if A feeds on C that feeds on B, and so on. Species may feed on only one (**monophages**), a few (**oligophages**), or many (**polyphages**) prey species.



Box 1 figure | Compartments, omnivory, clustering and nestedness.

a, b, Examples of food webs that are compartmented (**a**) and not compartmented (**b**). The web in **a** is clustered, and has within-chain omnivory (the ‘grain-fed hamburger in a bun’ mode). The web in **b** is not clustered, and its omnivory involves feeding on different food chains (the ‘fish and chips’ mode). The web in **c** has within-chain omnivory, but it is not clustered. **d, e**, Two mutualistic webs with identical numbers and distributions of trophic links, between pollinator species A to D and flower species 1 to 5. Interactions in web **d** (upper diagram) are not grouped into guilds, whereas in **e** (upper diagram) they are—as shown by the overlaps between the pollinators (lower diagrams).

Evolutionary patterns in diet specialization

Webs of specialists are less complex than those of generalists, so what determines whether a species specializes on few prey species or whether it takes many? The trade-offs between specialization and generalization are complex and multi-factorial^{4,39}. Monophages (species that eat just one other species) are probably more efficient consumers than are generalists—they usually have higher *per capita* ingestion and assimilation rates⁴⁰. Prey species also differ in the number of predators that feed on them. Experimental studies find that species consumed by multiple predators have lower predation rates than species with single predators, in contrast to the expected sum of the individual effect of each predator⁴¹. Combined, these results suggest that trophic specialists would have dynamically stronger interactions with individual prey species (and vice versa) than would more generalized species. We discuss the implications for food web dynamics below.

How are feeding interactions organized, given these patterns of relative specialization or generalization? Broad patterns are clear. For example, hummingbirds exploit different species of flowers than do insects. Within the class of insect-pollinated flowers, bees and butterflies feed on different sets of plant species⁴. Yet conventional wisdom suggests that such patterns should also apply within broad taxonomic groupings. Compelling examples⁴² of hummingbirds with long or strangely curved bills and the corresponding morphological specialization of the flowers they pollinate beg an obvious question: how ubiquitous are tight links between species of pollinator (or frugivore) and the flowers (or fruits) on which they feed?

The idea of reciprocal specializations reinforced by co-evolution seems so well entrenched that it is surprising that studies have not assessed its generality. The problem has probably been the exact measure of ‘guilds’ (groups of species exploiting the same resources by similar means, such as seed-eating birds) and the lack of a suitable null hypothesis against which to judge their improbability. Fortunately, there is an exact mathematical parallel to studies of the presence or absence of species on islands—a topic with a long, controversial history in ecology⁴³. Both problems involve a binary matrix where, under the null hypothesis, the row and column sums must remain the same as those observed in nature. Specifically, a given species must occur on a set total number of islands, whereas a given island will host a set total number of species. Likewise, a pollinator exploits a set total number of plant species and a plant offers nectar to a set total number of species. As with biogeographical questions, the computational problem involves generating null hypotheses that obey these constraints but are not artefactually close to the observed pattern. This issue is now solved⁴³, but no one has applied this technique to webs of mutualists. Nor, for that matter, has it been applied to sets of predators and their prey, though

Box 2 | Food web data limitations

Early studies^{1–3} analysed webs where the authors’ purpose generally was not to make comparisons between different webs. Important exceptions are from tree holes and other water-filled plant cavities⁶⁹ and plants, herbivores, their parasites, and hyperparasites²⁷—systems that continue to generate standard, often quantitative, methods to describe and compare webs^{23,70}. Responding to criticisms^{66,71,72} of early studies, recent efforts^{18,31,63} produced species-rich, linkage-dense webs.

Early studies also provided little or no information on the strength of trophic links. Following an appeal for more consistent data⁷², attempts to estimate such strengths find that they vary considerably⁵². Data on strengths generally involve only few links⁵², even when most of the system’s links are described⁷³. Identifying strong and weak links is difficult. Having a simple measure correlated with strength would be very helpful. Recent empirical⁵⁰ and theoretical⁷⁴ analyses find that predator-to-prey body size ratio predicts strength—the larger the ratio, the stronger the interaction.

there are abundant sources of such information—from fish and their prey⁴⁴, for example.

Mutualistic webs are strongly nested^{5,8}. Nested links constrain reciprocal specializations, but do not completely preclude them (see Box 1). Nonetheless, the ubiquity of nestedness is a surprise, showing as it does that specialized pollinators (for example) typically exploit flowers used by many other species. Moreover, the interactions are not symmetric: the plant species that is most important to a particular pollinator may receive more attention from other species that are not so dependent upon it⁴⁵.

Fragility and the paradox of complex interactions

Most differential equation models must be sufficiently 'simple' if the species are to persist at a locally stable equilibrium^{3,15,16,46}. Indeed, there is a striking coincidence in the simple, sparse patterns of links observed in webs and those that models predict^{15,16}. Sensibly parameterized models of real food webs tend to be locally stable, despite their complexity, suggesting that real webs have statistically special parameters and patterns of interactions^{47–50}. Natural history explanations pre-date dynamical models¹⁵ in suggesting that species that overlap too extensively in their diets cannot coexist. Importantly though, models do permit tightly clustered links and extensive omnivory in systems dominated by parasitoids³ and life-history omnivores²⁶. Because these classes of interaction dominate the

group of unusually clustered models, then perhaps this is a sufficient answer: such clusters are not prohibited by the dynamics as we understand them.

Another way of phrasing the problem of complex webs is to notice that the effect of one species on the density of another diminishes with their separation in the food web, measured by the distance connecting them⁵¹. The commonness of short paths in food webs^{6,7} suggests that disturbances spread rapidly throughout the food web. Certainly, species may not be close if the shortest links between them are weak or intermittent. As a species feeds on more prey species, or suffers the attention of more predators, the strength of the interactions may decrease. The 'many weak and few strong' pattern of interactions in quantified empirical food webs⁵² and webs of mutualists^{34,53} usually correlates with diet breadth. Polyphages (species eating many other species) and prey with many predators present weak interactions, whereas strong interactions correspond to monophages and prey with only a few predators. Intriguingly, Berlow⁵⁴ finds that in some systems the weak interactions have the greatest variance: they yield strong effects intermittently.

There is a long-recognized paradox: in models with random, but sensibly chosen, parameters, the probability of a locally stable equilibrium declines sharply as complexity increases⁴⁶. Yet, the ever-smaller fraction of models that are stable tend to be more resilient³. (Resilient in the narrow sense that the largest eigenvalue

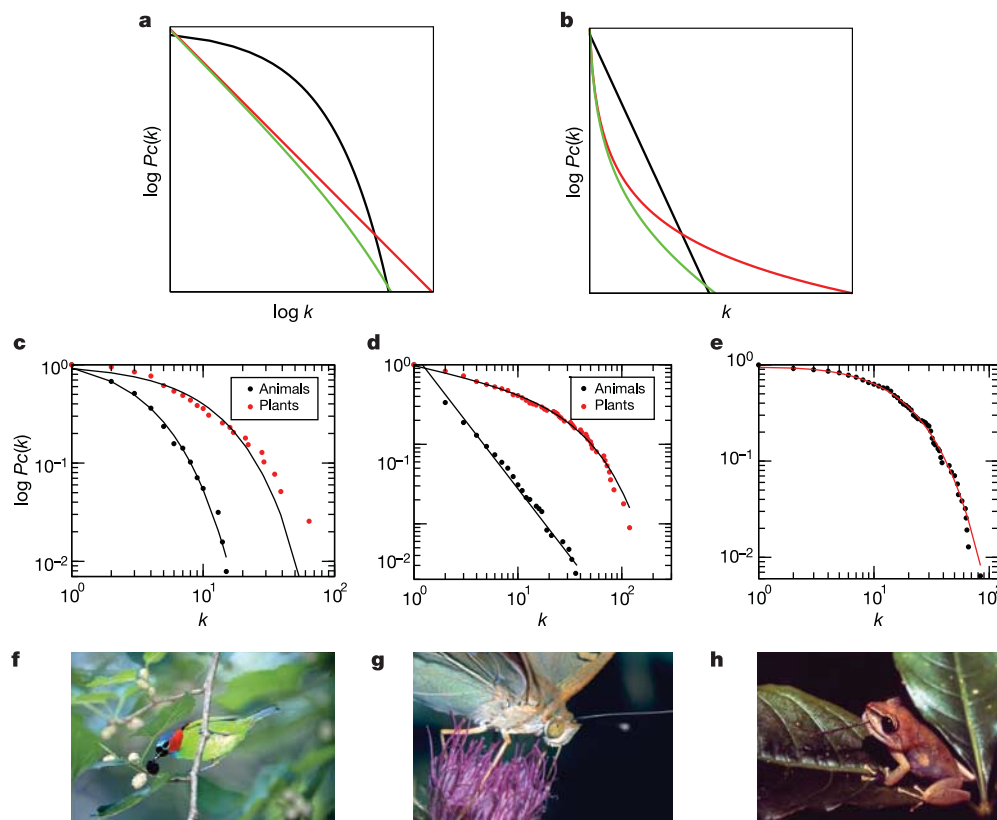


Figure 2 | Distribution of linkage density in ecological networks. **a–e**, The cumulative probabilities $P_c(k)$, for $\geq k$, where $P(k)$ is the probability a species has k links to other species, and is given by $P(k) \sim k^{-\gamma} e^{-k/\gamma}$ where $e^{-k/\gamma}$ introduces a cut-off at some characteristic scale γ . Panels **a** (log–log) and **b** (log–linear) show three different modelled networks. Black lines, single-scale networks; when γ is very small, the distribution has a fast decaying tail, typically exponential, $P(k) \sim e^{-k/\gamma}$. Green lines, truncated scale-free networks; these correspond to intermediate values of γ where the distribution has a power law regime followed by a sharp cut-off, with an exponential decay of the tail. Red lines, scale-free networks; for large values of γ the number of connections per species decays as a power law, $P(k) \sim k^{-\gamma}$, a function with a relatively 'fat tail'. **c–e**, Experimental data

(filled circles) and best fits (lines); **c**, a frugivore–plant web⁷⁶; **d**, a pollinator–plant web⁷⁷; and **e**, the food web from El Verde rainforest⁷⁸. In **c** and **d**, for red circles, k is the number of plants species visited by an animal, and for black circles, k is the number of pollinator species visiting each plant species. In **e**, we sum prey–predator links and predator–prey links for each species. Best fits to the data in **c–e** are as follows: **c**, animals, exponential, $P(k) = e^{-k/3.998}$; plants, truncated power law, $P(k) = k^{-0.013} e^{-k/11.22}$; **d**, animals, power law, $P(k) = k^{-1.512}$; plants, truncated power law, $P(k) = k^{-0.2822} e^{-k/42.55}$; **e**, exponential, $P(k) = e^{-k/8.861}$. **f–h**, Photographs of a frugivore–plant (**f**), insect–flower (**g**) and predator–prey (**h**) interaction of webs depicted in **c–e**, respectively.

in the linearized system will be more negative and disturbances will dissipate more quickly^{14,15}.) McCann *et al.*⁵⁵ developed this idea to argue the importance of weak linkages and their particular arrangement.

Indirect effects—those mediated by intermediate species—are often in the opposite direction to those expected on the basis of direct interactions alone^{15,51,56}. They are best known from experiments in the intertidal zones of sea-shores. Menge's review⁵⁷ of 23 of them is key. As species richness increased, a typical species interacted strongly with more species and through more indirect pathways.

Complex webs also pose a challenge when one considers severe changes such as the addition or removal of species. What will be the consequences of extinction rates that are now hundreds to thousands of times faster than normal⁵⁸? Consider the apparent paradox of two classic results: MacArthur's model⁵⁹ and Paine's experiment of removing a top predator from the inter-tidal zones⁶⁰. Dynamical models of species removals show that the more polyphagous the predators, the less the effect caused by removing one of its prey species (measured as the probability that the predator will be lost³). Thus, MacArthur's model is correct, but it is incomplete. Paine showed, and models confirm¹⁵ that removing the predators of polyphages eliminates many of the latter's prey species. Thus, whether simple or highly connected model food webs are robust to the loss of species depends entirely on whether one looks at top predators or plant species¹⁵. Moreover, changes in species composition may be large while changes in the biomass may be small, and vice versa¹⁵.

Into this already complex story come recent studies using strictly topological approaches (that is, without population dynamics) to actual webs, as against abstractions of them. When the most connected species are successively removed from the web, the web is not 'species deletion stable'³. Many species lose their only prey source (and so must become extinct in turn), and the web quickly breaks into many disconnected sub-webs^{61–64}. By contrast, when disasters extirpate species randomly, these webs are robust, showing both little fragmentation and few secondary extinctions^{61–64}. Perhaps well-connected species are those that are relatively abundant at their particular trophic level and so are unlikely to be lost.

In sum, we do not completely understand how species persist in complex webs when disturbances propagate quickly and apparently strongly. More realistic models that incorporate the relationships between link strength, abundance and trophic generalization may provide the answers.

Conclusions

Our knowledge of the structure of ecological networks is still incomplete in important areas that include compartments and reciprocal specialization. Species are closer than previously thought, and many webs have very dense clusters of linkages. The relationships between body size, abundance, trophic specialization and nested diets (both across and within trophic levels) are not only interesting in themselves^{9,35,38}; they may constrain web dynamics in theoretically unexplored ways. Such constraints may be essential to explain the persistence of species in a constantly changing world, and the tolerance of current ecosystems to natural gains and losses of species as well as their vulnerability to unnaturally inflated extinction rates.

- Pimm, S. L., Lawton, J. H. & Cohen, J. E. Food web patterns and their consequences. *Nature* **350**, 669–674 (1991).
- Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* (Springer, Berlin, 1990).
- Pimm, S. L. *Food Webs* (Univ. Chicago Press, Chicago, 2003).
- Thompson, J. N. *The Coevolutionary Process* (Univ. Chicago Press, Chicago, 1994).
- Jordano, P., Bascompte, J. & Olesen, J. M. in *Plant-Pollinator Interactions: From Specialization to Generalization* (eds Waser, N. M. & Ollerton, J.) (Univ. Chicago Press, Chicago, 2005).
- Montoya, J. M. & Solé, R. V. Small world patterns in food webs. *J. Theor. Biol.* **214**, 405–412 (2002).
- Williams, R. J., Berlow, E. L., Dunne, J. A., Barabási, A.-L. & Martinez, N. D. Two degrees of separation in complex food webs. *Proc. Natl Acad. Sci. USA* **99**, 12913–12916 (2002).
- Bascompte, J., Jordano, P., Melian, C. J. & Olesen, J. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
- Woodward, G. *et al.* Body-size in ecological networks. *Trends Ecol. Evol.* **20**, 402–409 (2005).
- Dunne, J. E., Williams, R. J. & Martinez, N. D. Food web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12917–12922 (2002).
- Strogatz, S. Exploring complex networks. *Nature* **410**, 268–276 (2001).
- Álbert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Solé, R. V., Ferrer, R., Montoya, J. M. & Valverde, S. Selection, tinkering and emergence in complex networks. *Complexity* **8**, 20–33 (2002).
- Pimm, S. L. The complexity and stability of ecosystems. *Nature* **307**, 321–326 (1984).
- Pimm, S. L. *The Balance of Nature? Ecological Issues in the Conservation of Species and Communities* (Univ. Chicago Press, Chicago, 1991).
- Solé, R. V. & Bascompte, J. *Self-organization in Complex Ecosystems* (Princeton Univ. Press, Princeton, 2006).
- Warren, P. H. in *Aspects of the Genesis and Maintenance of Biological Diversity* (eds Hochberg, M. E. *et al.*) 142–161 (Oxford Univ. Press, Oxford, 1996).
- Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
- Williams, R. J. & Martinez, N. D. Limits to trophic levels and omnivory in complex food webs: theory and data. *Am. Nat.* **163**, 458–468 (2004).
- Pimm, S. L. & Lawton, J. H. Are food webs compartmented? *J. Anim. Ecol.* **49**, 879–898 (1980).
- Thomas, J. D. Mutualistic interactions in freshwater modular systems with molluscan components. *Adv. Ecol. Res.* **20**, 125–178 (1990).
- Fonseca, C. R. & Ganade, G. Asymmetries, compartments and null interactions in an Amazonian ant–plant community. *J. Anim. Ecol.* **65**, 339–347 (1996).
- Memmott, J., Godfray, H. C. J. & Gould, I. D. The structure of a tropical host–parasitoid community. *J. Anim. Ecol.* **63**, 521–540 (1994).
- Raffaelli, D. G. & Hall, S. J. Compartments and predation in estuarine food web. *J. Anim. Ecol.* **61**, 551–560 (1992).
- Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E. & Taylor, W. W. Compartments revealed in food-web structure. *Nature* **426**, 282–285 (2003).
- Pimm, S. L. & Rice, J. A. The dynamics of multispecies, multi-life-stage models of aquatic food webs. *Theor. Popul. Biol.* **32**, 303–325 (1987).
- Askew, R. R. On the biology of inhabitants of oak galls of Cynipidae (Hymenoptera) in Britain. *Trans. Soc. Br. Entomol.* **14**, 237–269 (1961).
- Strong, D. R., Lawton, J. H. & Southwood, R. *Insects on Plants* (Harvard Univ. Press, New York, 1984).
- Amaral, L. A. N., Scala, A., Barthélemy, M. & Stanley, H. E. Classes of behavior of small-world networks. *Proc. Natl Acad. Sci. USA* **97**, 11149–11152 (2000).
- Jordano, P., Bascompte, J. & Olesen, J. M. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* **6**, 69–81 (2003).
- Montoya, J. M. & Solé, R. V. Topological properties of food webs: from real data to community assembly models. *Oikos* **102**, 614–622 (2003).
- Stouffer, D. B., Camacho, J., Guimera, R., Ng, C. A. & Nunes Amaral, L. A. Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86**, 1301–1311 (2005).
- Garlaschelli, D., Caldarelli, G. & Pietronero, L. Universal scaling relations in food webs. *Nature* **423**, 165–168 (2003).
- Vázquez, D. P. & Aizen, M. A. Null model analyses of specialization in plant–pollinator interactions. *Ecology* **84**, 2493–2501 (2003).
- Cohen, J. E., Jonsson, T. & Carpenter, S. R. Ecological community description using the food web, species abundance, and body size. *Proc. Natl Acad. Sci. USA* **100**, 1781–1786 (2003).
- Sugihara, G., Bersier, L.-F., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100**, 5246–5251 (2003).
- Cattin, M. F., Bersier, L.-F., Banasek-Richter, C., Baltensperger, M. & Gabriel, J.-P. Phylogenetic constraints and adaptation explain food-webs structure. *Nature* **427**, 835–839 (2004).
- Jonsson, T., Cohen, J. E. & Carpenter, S. R. Food webs, body size and species abundance in ecological community description. *Adv. Ecol. Res.* **36**, 1–84 (2005).
- Hawkins, B. A. *Pattern and Process in Host-Parasitoid Communities* (Cambridge Univ. Press, Cambridge, UK, 1994).
- Rana, J. S., Dixon, A. F. G. & Jarosik, V. Costs and benefits of prey specialization in a generalist insect predator. *J. Anim. Ecol.* **71**, 15–22 (2002).
- Sih, A., Englund, G. & Wooster, D. Emergent impacts of multiple predators on prey. *Trends Ecol. Evol.* **13**, 350–355 (1998).
- Temeles, E. J. & Kress, W. J. Adaptation in a plant–hummingbird association. *Science* **300**, 630–633 (2003).
- Sanderson, J. G. Testing ecological patterns. *Am. Sci.* **80**, 332–339 (2000).
- Froese, R. & Pauly, D. (eds) *FishBase*. <<http://www.fishbase.org>> (2005).
- Bascompte, J., Jordano, P. & Olesen, J. M. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science* **312**, 431–433 (2006).

46. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, Princeton, 1974).
47. Yodzis, P. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *J. Anim. Ecol.* **67**, 635–658 (1998).
48. McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
49. Neutel, A.-M., Heesterbeek, J. A. P. & de Ruiter, P. C. Stability in real food webs: weak links in long loops. *Science* **296**, 1120–1223 (2002).
50. Emmerson, M. & Raffaelli, D. Body size, patterns of interaction strength and the stability of a real food web. *J. Anim. Ecol.* **73**, 399–409 (2004).
51. Abrams, P., Menge, B. A., Mittelbach, G. G., Spiller, D. & Yodzis, P. in *Food Webs: Integration of Patterns and Dynamics* (eds Polis, G. A. & Winemiller, K. O.) 371–395 (Chapman & Hall, New York, 1996).
52. Berlow, E. *et al.* Interaction strengths in food webs: issues and opportunities. *J. Anim. Ecol.* **73**, 585–598 (2004).
53. Jordano, P. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *Am. Nat.* **129**, 657–677 (1987).
54. Berlow, E. L. Strong effects of weak interactions. *Nature* **398**, 330–334 (1999).
55. McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
56. Wootton, J. T. Indirect effects in complex ecosystems: recent progress and future challenges. *J. Sea Res.* **48**, 157–172 (2002).
57. Menge, B. A. Indirect effects in marine rocky intertidal interaction webs: patterns and importance. *Ecol. Monogr.* **65**, 21–74 (1995).
58. Pimm, S. L. *World According to Pimm: a Scientist Audits the Earth* (McGraw-Hill, New York, 2001).
59. MacArthur, R. H. Fluctuations of animal populations and a measure of community stability. *Ecology* **36**, 533–536 (1955).
60. Paine, R. T. A note on trophic complexity and community stability. *Am. Nat.* **103**, 91–93 (1969).
61. Álbért, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
62. Solé, R. V. & Montoya, J. M. Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B* **268**, 2039–2045 (2001).
63. Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
64. Memmot, J., Waser, N. M. & Price, M. V. Tolerance of pollination networks to species extinctions. *Proc. R. Soc. Lond. B* **271**, 2605–2611 (2004).
65. Watts, D. J. & Strogatz, S. H. Collective dynamics in ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
66. Hall, S. J. & Raffaelli, D. Food web patterns: lessons from a species-rich web. *J. Anim. Ecol.* **60**, 823–842 (1991).
67. Polis, G. A. & Strong, D. R. Food web complexity and community dynamics. *Am. Nat.* **147**, 813–846 (1996).
68. Root, R. B. The niche exploitation pattern of the blue-gray gnatcatcher. *Ecol. Monogr.* **37**, 317–350 (1967).
69. Kitching, R. L. in *Phytotelmata: Terrestrial Plants as Hosts of Aquatic Insect Communities* (eds Frank, H. & Lounibos, P.) 205–222 (Plexus, New Jersey, 1983).
70. Kitching, R. L. *Food Webs and Container Habitats* (Cambridge Univ. Press, Cambridge, UK, 2000).
71. Polis, G. A. Complex desert food webs: an empirical critique of food web theory. *Am. Nat.* **138**, 123–155 (1991).
72. Cohen, J. E. *et al.* Improving food webs. *Ecology* **74**, 252–258 (1993).
73. Raffaelli, D. G. & Hall, S. J. in *Food Webs: Integration of Patterns and Dynamics* (eds Polis, G. A. & Winemiller, K. O.) 185–191 (Chapman & Hall, New York, 1996).
74. Emmerson, M. E., Montoya, J. M. & Woodward, G. in *Dynamic Food Webs* (eds de Ruiter, P. C., Wolters, V. & Moore, J. C.) 179–197 (Academic, San Diego, 2005).
75. Melian, C. J. & Bascompte, J. Food web cohesion. *Ecology* **85**, 352–358 (2004).
76. Wheelwright, N. T., Haber, W. A., Murray, K. G. & Guindon, C. Tropical fruit eating birds and their food plants: a survey of a Costa Rican lower montane forest. *Biotropica* **16**, 173–192 (1984).
77. Inoue, T., Kato, M., Kakutani, T., Suka, T. & Itino, T. Insect–flower relationship in the temperate deciduous forest of Kibune, Kyoto: an overview of the flowering phenology and the seasonal pattern of insect visits. *Contrib. Biol. Lab. Kyoto Univ.* **27**, 377–463 (1990).
78. Reagan, D. P. & Waide, R. B. *The Food Web of a Tropical Rain Forest* (Univ. Chicago Press, Chicago, 1996).

Acknowledgements We thank J. Dunne, D. Raffaelli, M. Emmerson and G. Woodward for comments; P. Jordano and R. Waide for the pictures and data for Fig. 2; and S. Valverde for assistance with the figures. This work was supported by an NERC Fellowship (J.M.M.) and the Santa Fe Institute (grant to R.V.S.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to S.L.P. (stuartpimm@aol.com).

Structural asymmetry and the stability of diverse food webs

Neil Rooney¹, Kevin McCann¹, Gabriel Gellner¹ & John C. Moore²

Untangling the influence of human activities on food-web stability and persistence is complex given the large numbers of species and overwhelming number of interactions within ecosystems. Although biodiversity has been associated with stability, the actual structures and processes that confer stability to diverse food webs remain largely unknown. Here we show that real food webs are structured such that top predators act as couplers of distinct energy channels that differ in both productivity and turnover rate. Our theoretical analysis shows that coupled fast and slow channels convey both local and non-local stability to food webs. Alarming, the same human actions that have been implicated in the loss of biodiversity also directly erode the very structures and processes that we show to confer stability on food webs.

May^{1,2} ignited one of the most enduring debates in ecology when he proposed that there is no *a priori* mathematical reason to believe that greater species-diversity begets enhanced community stability, and further challenged ecologists to “...elucidate the devious strategies which make for stability in enduring natural systems”. If increased diversity does not necessarily result in greater stability, why do diverse food webs seem to be more stable than depauperate ones? May noted that if the interactions among species within his models were arranged into ‘blocks’ rather than at random, stability was more likely, suggesting that the pattern of interactions in real food webs might impart stability in nature. This intuition has been supported by studies revealing that models structured from biologically plausible interaction strengths generate more stability than their randomly generated counterparts^{3–5}. Although increased stability in real food webs has thus been documented, ecologists have yet to identify the ecological basis for these structures and the mechanisms that impart stability in nature.

Ecologists have pointed to the potential for both sub-assemblages of species or compartments within food webs and the concomitant structure of interaction strengths among species to have significant roles in the persistence of complex natural systems^{1,2,4,6}. Whereas recent analyses have shown promise in analysing patterns of interaction strengths in nature to reveal the presence of compartments in large food webs⁷, ecologists have vacillated on the roles of compartments and interaction strengths in nature^{8–10}. Recent theoretical work, however, suggests that both may indeed have critical roles in ecosystem stability^{11,12}. Despite theoretical^{1,13,14} and empirical^{15–18} advances, the complex tangle of interactions found in real food webs has made it difficult to link the patterns of interaction strengths in nature and the mechanisms that impart stability to food webs^{19,20}.

A primary goal for ecologists, therefore, should be to identify the overarching structures and processes in food webs that impart stability and persistence. Along these lines, soil ecologists have argued the existence of compartments in the form of fast and slow energy channels, ultimately coupled by mobile higher-order consumers in soil food webs^{4,21–23}. Here, fast and slow channels refer to the turnover rates, or the production:biomass ratios, which are intimately related to energy or biomass flux definitions of interaction strength^{6,23} (see Box 1). Turnover rates can be seen as a combination of interaction

strength and the life history attributes of species or trophospecies—for example, birth rates, death rates and energetic efficiencies. Thus, if soil ecologists’ suggestions are correct, fast channels are on average comprised of strong interactions and slow channels are comprised of weak interactions. This simple connection between turnover rate and interaction strength allows us to examine food-web turnover rates that emanate from distinct basal resources (source webs *sensu* Cohen²⁴). If patterns of fast and slow energy channels exist in food webs, then in a sense we effectively have located strong and weak interaction chains. Here we expand on this suggestion of soil ecologists^{4,12–14} by rigorously assessing the existence of distinct energy channels in a variety of ecosystems. We then assess if there is, in fact, any tendency towards characteristic turnover rates within channels. Finally, we synthesize our empirical results within recent theory and in doing so show that coupled fast and slow energy channels confer both potent local and non-local stability to food webs.

Empirical results

On the existence of energy channels. We began our empirical analysis with the identification of basal resources within aquatic and terrestrial ecosystems from published food-web data. For aquatic food webs, we analysed data from Chesapeake Bay^{25,26}, the Chilean upwelling²⁷, the Cantabrian Sea shelf²⁸, and the Bering Sea²⁹. Terrestrial food webs included in the study were a North American grassland (shortgrass steppe, SGS-LTER Nunn, Colorado)³⁰, the Lovinkhoeve Experimental Farm (Integrated Management) in The Netherlands³¹, moist acidic tundra (Arctic-LTER, Toolik Lake, Alaska)³², and a European Scots pine forest (Werkerom, The Netherlands)³³. In our examples, aquatic systems consisted of resource compartments based on phytoplankton and detritus. In three of the four detrital soil food webs studied, bacteria and fungi were the basal resources, whereas in the fourth terrestrial example, fungi and detritus were the basal resources (see Supplementary Notes for data sources). On the basis of feeding interactions, we calculated the per cent of carbon derived from each basal resource by any given consumer, and the trophic position of each food-web member (see Supplementary Methods for details). Figure 1 shows two examples of the general pattern that emerged from our analysis (see Supplementary Fig. 1a, b for six further examples). Lower-order consumers tended to derive the bulk

¹Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. ²School of Biological Sciences, University of Northern Colorado, Greeley, Colorado 80639, USA.

of their carbon from one of the two resources, providing the basis for resource compartments or energy channels. As trophic level increased, so did the tendency to derive carbon from both source webs. Notably, analogous patterns of predatory fish coupling benthic and pelagic pathways have also been detected using stable isotope analysis in lake food webs³⁴. Given the broad range of systems analysed, the consistency of the pattern among food webs is remarkable and, to our knowledge, novel.

On productivity and turnover rates. As strong and weak channels map to fast and slow turnover rates within compartments, we calculated or took from the literature production and biomass values for the channels (defined as groups that derived >60% of their carbon from a given basal resource). Our analysis substantiates previously suspected patterns of weak and strong energy flux through channels. Table 1 shows that for the example food webs presented in Fig. 1, the amount of production entering aquatic food webs is consistently higher in the pelagic channel. Similarly, within the example terrestrial web, the consumers of bacteria dominate the carbon influx. Strikingly similar patterns of asymmetrical energy flux into food webs are also seen in the remaining six food webs

(Supplementary Table 1). The important point is that carbon influx into these food webs through primary consumers is not symmetrical, as one channel dominates.

Moving up trophic levels in the food webs, the production:biomass ratios within the pelagic channels are consistently higher than the benthic channels for aquatic systems (Table 1). This pattern reflects ecological and life history characteristics of the consumers within each channel. For example, pelagic primary consumers (zooplankton) are smaller and have much shorter generation times compared with their benthic primary consumer counterparts. By and large, bacterial channels showed higher production:biomass ratios than fungal channels in soil systems dominated by non-woody plants, such as the North American grassland.

The empirical analysis has detected remarkable similarities in food-web structure across a range of ecosystems. Food webs are constructed such that energy channels based on different basal resources are coupled by higher-order predators. Across systems, channels within food webs are consistently asymmetrical in both the amount of energy entering through basal resources (annual production) and the turnover rates of the energy channels (production:biomass ratios).

Box 1 | The relationship between turnover rate (P:B ratio) and interaction strength

Let us start with a generic consumer model where the consumer, C , feeds on resources, R_i , and is fed on by predators, P_j :

$$\frac{dC}{dt} = e \sum_{i \in R} CF_{CR_i} - mC - \sum_{j \in P} CF_{P_j C} \quad (1)$$

where F_{ji} is the functional response of consumer j on resource i , m is the rate of mortality, and e is the conversion efficiency. Assuming equilibrium conditions (that is, flux into C equals flux out of C), then consumer biomass turnover rates (ν_C) can be calculated from either the flux in or the flux out:

Flux in

$$\nu_C = \frac{e \sum_{i \in R} CF_{CR_i}}{C} \quad (2)$$

Flux out

$$\nu_C = \frac{\sum_{j \in P} CF_{P_j C} + mC}{C} \quad (3)$$

Using energy flow as a metric for interaction strength^{23,31}, the per capita interaction strength (IS) between any resource and its consumer can be defined as:

$$IS'_{CR} = -\frac{CF_{CR}}{C} \quad IS'_{RC} = \frac{eCF_{CR}}{R} \quad (4)$$

Similarly, the total interaction strength can be defined as:

$$IS_{CR} = -CF_{CR} \quad IS_{RC} = eCF_{CR} \quad (5)$$

Given estimates of biomass, then the magnitudes of equations (2) to (5) can easily be estimated.

Thus, the relationship between consumer biomass turnover rates (ν_C) and interaction strengths (IS) can be expressed as:

Flux in

$$\nu_C = \frac{e \sum_{i \in R} CF_{CR_i}}{C} = \frac{IS_{R_1 C} + IS_{R_2 C} + \dots}{C} \quad (6)$$

Flux out

$$\nu_C = \frac{\sum_{j \in P} CF_{P_j C} + mC}{C} = \frac{-IS_{P_1 C} - IS_{P_2 C} - \dots + m}{C} \quad (7)$$

It is clear that the speed of energy flow for a consumer ultimately depends on a number of traits of an organism. The speed of energy flow (ν) can thus be envisioned as a general property of a food web (interaction strength is embedded within this property). Thus, a fast flow is, on average, composed of stronger interactions than a slow flow. Furthermore, fast pathways (for example, up a food chain), on average, are composed of stronger interactions than slow pathways.

Theoretical synthesis

Theoretical explorations of compartmented food webs have suggested that system dynamics can be stabilized by both higher-order predators coupling food webs in space, and the presence of weak and strong channels^{10,11}. These results resonate with our empirical observations, so we sought to integrate them into a food-web model to analyse their influence on food-web stability. Our model consists of a top predator that couples two consumer-resource chains that exist in different habitats. The resources have access to exclusive nutrient pools, and also compete for a common pool of nutrients (see Supplementary Fig. 2 for model details and Supplementary Table 2 for parameter values). Within the model, we varied two central factors that mirrored our empirical observations. To reflect the observed differences in production:biomass ratios within channels, we simultaneously varied the attack rates of both predator on consumer and consumer on resource (attack rate ratio). Given the constant attack rates of one channel (constant channel), we varied the attack rates of the second channel (modified channel) by factors ranging from 0.5 to 2.5, thus forcing the modified channel to be both slower and faster than the constant channel. In preliminary model runs, we also modified channel speed using a variety of different parameters (for example, mortality (m)) without qualitatively changing the results that follow. To parallel the observed differences in basal production between energy channels, we independently varied the proportion of energy (p) that entered the food web through the resources of the modified channel. This metric essentially forced food-web production to be dominated by either the constant channel or the modified channel (see Supplementary Fig. 1 and Table 1 for further details).

Equilibrium results. First we considered the most straightforward scenario, where equal amounts of energy enter the food web through the two basal resources ($p = 50\%$) and then varied the attack rate ratio (Fig. 2a, solid line). The least stable food web configuration occurs when the attack rate ratio is 1:1, a similar pattern found in previously published work¹¹. This symmetrical arrangement of energy flow essentially creates an unstructured food web, wherein the dynamics of the two channels are synchronized. Increasing or decreasing the attack rate ratio (analogous to increasing or decreasing the production:biomass ratio in the modified channel) confers local stability, as measured by the most negative eigenvalue (Fig. 2a). Stated biologically, we see that equilibrium stability is enhanced as energy flow between the channels becomes asymmetric. This change in energy flow is reflected in the dynamics of the consumers from the two different energy channels. Figure 2b shows that the correlation between consumers following a small perturbation decreases on

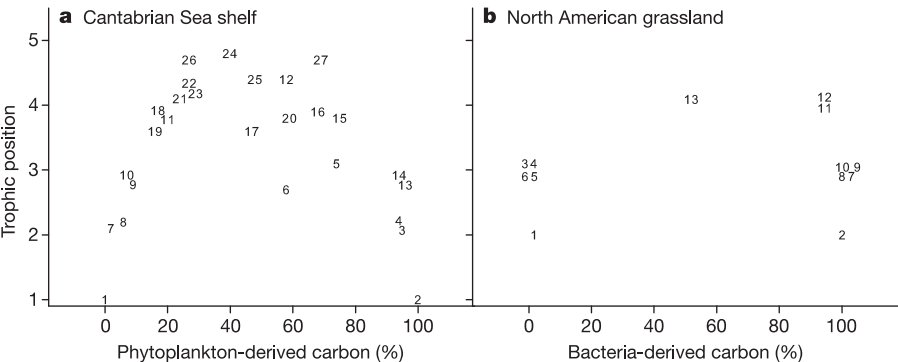


Figure 1 | Food webs are structured such that top predators couple distinct energy channels.
a, b, Food-web representations based on estimations of the percentage of carbon derived from basal resources for the Cantabrian Sea shelf (**a**) and the North American grassland (**b**). Trophic position (based on feeding interactions) is shown on the y-axis. Taxonomic group names corresponding to the numbers can be found in Supplementary Table 4. Labels have been offset to allow for the identification of taxonomic groups.

either side of the symmetrical, or unstructured, case (when the attack rate ratio = 1.0 and $p = 50\%$). This implies that the two chains start to vary out of phase as the attack rate ratio is skewed (Fig. 2b, solid line). As recent theory has suggested, this asynchrony in prey (that is, the consumers) translates into a more stable prey base for the predator^{12,35} and so can act as an important stabilizing mechanism in nature.

Mirroring our empirical results, we next asked how simultaneously varying both p and the attack rate ratio modified these results. Figure 2a (dashed line) plots the most stable configurations observed for any given attack rate ratio across the entire range in p . The local stability is greatly enhanced when we allow both factors to change (Fig. 2a), demonstrating that the most stable food web configurations occur when both attack rates and production (p) are skewed.

Increased stability of these food webs is reflected in lower correlation coefficients (greater asynchrony) between the consumer populations (Fig. 2b, dashed line). To unfold this pattern more clearly, we plotted the percentage of energy that flowed through the faster channel (remembering that the modified channel is fastest above an attack rate ratio of 1:1, and the constant channel is the faster below an attack rate ratio of 1:1) at the most stable scenario for each attack rate ratio (Fig. 2c). The first clear pattern that emerges is that food webs are always most stable when the majority of energy flows through the fast channel. Further, as the food webs become more asymmetric with respect to the attack rate ratio, more energy must flow through the fast channel to realize maximum stability.

Transient non-equilibrium results. Ecological systems are inherently variable and so it is worthwhile to consider how mechanisms that drive equilibrium stability act away from the equilibrium (for example, after a large perturbation). We allowed our model systems to reach equilibrium and then removed 10% of the top predator as a perturbation to the food web. To understand the dynamics within a

transient non-equilibrium context, we looked at two dynamical properties that confer resilience and stability to food webs. First, we determined the rate at which the predator population densities return to the equilibrium value immediately following a perturbation ('global return speed', Fig. 3a). Second, we determined the degree to which the predator population overshoots its equilibrium during the transient phase ('overshoot', Fig. 3a), as a measure of the potential for compensatory responses to dampen food-web dynamics near the equilibrium. These two attributes are important, as often dynamic systems that return to equilibrium rapidly also tend to overshoot the equilibrium significantly. Using these metrics, we constructed a measure of non-local stability ('transient non-equilibrium stability', Fig. 3a) defined as global return speed divided by overshoot. Defined this way, stability away from the equilibrium is largest when global return speed is large and overshoot is simultaneously small. As illustrated in Fig. 3b, the regions of greatest local stability (most negative dominant eigenvalue, Fig. 2a) also have high degrees of transient non-equilibrium stability. Although we do not explore the influence of environmental fluctuations on the stability of our food web configuration, we argue that the very same factors that confer stability away from the system attractor in our example will do so in nature. Both the ability to respond quickly to changes and the capacity to dampen potentially destabilizing oscillatory fluctuations will confer overall stability to food webs in the face of fluctuating environmental conditions.

Summary

The phenomenon of population asynchrony resulting in community stability has been postulated in theoretical analyses^{12,35} and observed in both experimental³⁶ and empirical³⁷ studies. Here we demonstrate that common empirical food-web structures have the potential to readily drive such a differential response. Encouragingly, this suggests that the weak interaction effect, heretofore demonstrated at the

Table 1 | Energy flow in a marine and a terrestrial food web

Cantabrian Sea shelf	Trophic level	Benthic channel			Pelagic channel		
		Production (g C m ⁻² yr ⁻¹)	Biomass (g C m ⁻²)	P:B ratio	Production (g C m ⁻² yr ⁻¹)	Biomass (g C m ⁻²)	P:B ratio
	2 +	130.38	35.15	3.71	537.30	22.68	23.69
	3 +	14.76	20.47	0.72	72.79	29.76	2.45
	4 +	1.10	2.49	0.44	0.31	0.38	0.82
North American grassland	Trophic level	Fungal channel			Bacterial channel		
		Production (g C m ⁻² yr ⁻¹)	Biomass (g C m ⁻²)	P:B ratio	Production (g C m ⁻² yr ⁻¹)	Biomass (g C m ⁻²)	P:B ratio
	2 +	11.00	6.30	1.75	54.55	30.40	1.8
	3 +	0.64	0.35	1.83	5.19	1.14	4.5
	4 +	–	–	–	0.032	0.016	2.0

Total production, biomass and production:biomass (P:B) ratios of species grouped by trophic position and basal energy resource for the two example food webs illustrated in Fig. 1. Values for the remaining six food webs can be found in Supplementary Information.

consumer–resource level⁶, may be scaled to whole ecosystems if we recognize the presence of fast and slow energy channels coupled by higher-order predators (that is, the fast and slow channels can be envisioned as coupled strong and weak interaction chains). It is important to point out that this stabilizing result depends on the rapid behavioural response of the top predator to changing densities in the different channels³⁸. As one channel increases and the other decreases (that is, the channels have asynchronous dynamics) the predator moves to regulate the increasing channel and in doing so frees the decreasing channel from strong predatory pressure. It is therefore the rapid predatory switching behaviour that balances such asynchrony. Although here we use a preference-based functional response in our model, optimal foraging models will give the same

qualitative results as they too respond to the asynchrony in a way that balances such asynchronous dynamics^{39,40}.

Notably, our results also suggest that a food-web architecture based on multiple asymmetric energy channels provides ecosystems with a potent mechanism for responding to large perturbations (that is, dynamics far from the equilibrium). The asymmetry of energy flux in coupled channels provides top predators with prey bases that show characteristic and complementary dynamics. In the face of such large perturbations, a fast channel allows for the rapid recovery of predator populations, but would almost certainly result in overshoot and runaway consumption dynamics. The presence of a slow channel, however, enhances compensatory responses near the system attractor. These complementary functions produce a rapid yet stable recovery from a perturbation.

In summary, our results suggest that the stability of complex ecosystems depends critically on the maintenance of the heterogeneity of distinct energy channels, their differential dynamic properties (that is, differential productivity and turnover), and the mobile consumers that couple these distinct channels. A corollary to this general result is twofold: (1) any perturbations that destroys this heterogeneity (either through synchronizing or removing channels) ought to destabilize these systems relative to their heterogeneous counterparts; and, (2) removal of mobile higher-order consumers that couple these distinct zones ought to destabilize these systems relative to their coupled counterparts.

It is important to note that the same traits we identify as critical ecosystem structures are currently being undermined by human

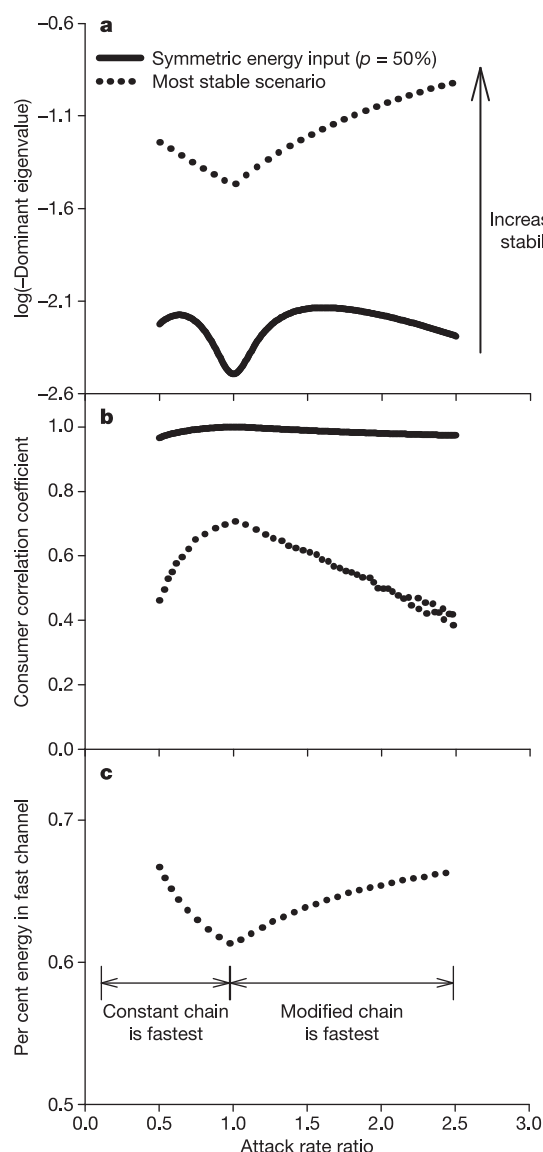


Figure 2 | Differing productivities and turnover rates between coupled energy channels result in increased local food-web stability and decreased consumer population synchrony. **a–c**, The relationship between attack rate ratio and: **a**, the local stability of the food webs ($\log(-\text{dominant eigenvalue})$); solid line represents food-web stability when $p = 50\%$ and the dotted line is the most stable possible state given any p ; **b**, the correlation coefficient of the two consumer populations during the transient phase of the food-web dynamics (legend same as in **a**); and **c**, the proportion of energy flowing through the faster channel that gives rise to the greatest stability as shown in **a**. See Supplementary Fig. 1 for model equations and parameter values used in the food-web model.

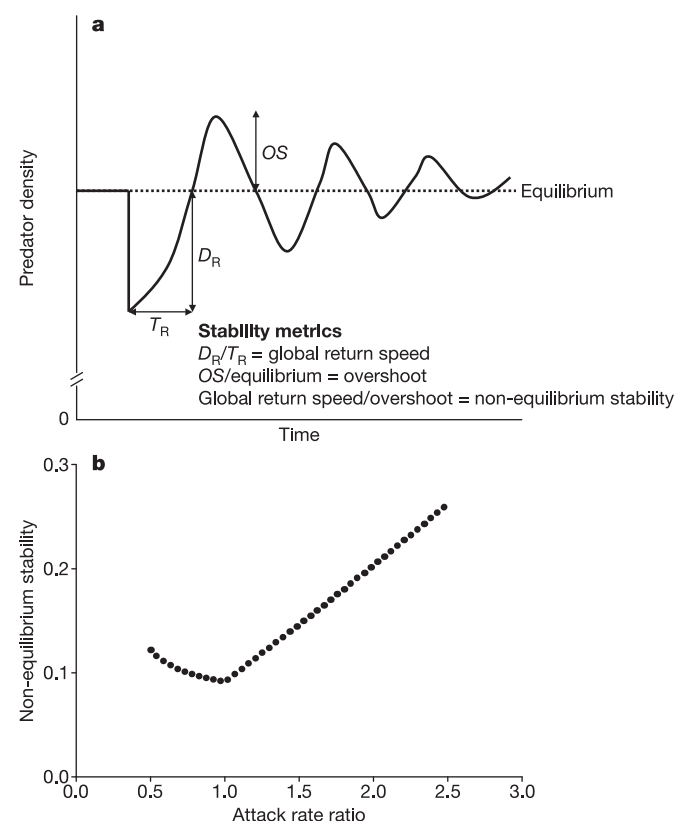


Figure 3 | Coupled fast and slow energy channels allow for non-local food-web stability. **a**, A schematic diagram assessing transient non-equilibrium stability using post-perturbation predator densities. After a perturbation, D_R is the increase in density required to reach equilibrium, T_R is the time required to reach equilibrium, and OS is the overshoot in predator density. The metrics measured quantify the rate at which food webs react to perturbations and the degree to which overshoot is dampened by food-web interactions. **b**, The relationship between attack rate ratio and the transient non-equilibrium stability of the food webs.

activities. For example, the literature is replete with documented losses of top predators (Supplementary Table 3a), and, in cases where ensuing food-web dynamics have been documented, diversity has abruptly and dramatically declined (termed 'ecological meltdown' by Terborgh and colleagues⁴¹). In addition, human activities have also frequently resulted in dramatic changes to both the flux of energy into food webs, and the process rates within energy channels (Supplementary Table 3b). Nutrient loading to ecosystems often drives almost complete dominance by a specific energy channel such that we effectively have homogenized production in such systems. Ecosystems can often resist such loading up to a point whereupon a catastrophic shift occurs, resulting in both the loss of diversity and stability of these ecosystems⁴². Thus, it seems that human actions are not only reducing the diversity of our natural ecosystems, but more importantly they are probably eroding the very structures that confer stability to food webs in nature.

METHODS

We used food-web data that contained consumption rates for all consumer–resource interactions. Within the soil food webs, only carbon derived from detrital sources was incorporated. Interactions within food webs were all converted to flux rates of $\text{g C m}^{-2} \text{yr}^{-1}$. Producers and detritus were assigned trophic positions of 1 and, where not reported directly, higher-order consumer trophic positions (TP_C) were calculated as:

$$TP_C = 1 + \sum_i^n P_C \times TP_R$$

where n is the number of resources consumed by the consumer, P_C is the proportion of the consumers' diet accounted for by a resource, and TP_R is the trophic position of the resource. In a similar manner, the proportion of carbon derived from basal resources (% BR_C) was calculated as:

$$\%BR_C = \sum_i^n P_C \times \%BR_R$$

where n is the number of resources consumed by the consumer, P_C is the proportion of the consumers' diet accounted for by a resource, and % BR_R is the proportion of carbon derived from the basal resource in the resource being consumed. In some cases, for calculating both trophic position and the proportion of carbon derived from a basal resource, loops within food webs necessitated iterations of the calculations.

In cases where biomass production was not reported directly, production values were estimated by assuming equilibrium conditions using:

$$P_C = e \sum_{i \in R} CF_{CR}$$

where R is the basal resource, C is the consumer eating R , F_{CR} is the functional response of the consumer on the resource, and e is the conversion efficiency.

Received 31 March; accepted 10 May 2006.

- May, R. M. Stability in multi-species community models. *Math. Biosci.* **12**, 59–79 (1971).
- May, R. M. Stability and complexity in model ecosystems (Princeton Univ. Press, Princeton, 1973).
- Yodzis, P. The stability of real ecosystems. *Nature* **289**, 674–676 (1981).
- de Ruiter, P. C., Neutel, A. M. & Moore, J. C. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* **269**, 1257–1260 (1995).
- Neutel, A. M., Heesterbeek, J. A. P. & de Ruiter, P. C. Stability in real food webs: weak links in long loops. *Science* **296**, 1120–1123 (2002).
- McCann, K. S., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
- Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E. & Taylor, W. W. Compartments revealed in food web structure. *Nature* **426**, 282–285 (2003).
- Pimm, S. L. The structure of food webs. *Theor. Popul. Biol.* **16**, 144–158 (1979).
- Pimm, S. L. & Lawton, J. H. Are food webs compartmented? *J. Anim. Ecol.* **49**, 879–898 (1980).
- Ives, A. R. & Hughes, J. B. General relationships between species diversity and stability in competitive systems. *Am. Nat.* **159**, 388–395 (2002).
- Post, D. M., Conners, M. E. & Goldberg, D. S. Prey preference by a top predator and the stability of linked food chains. *Ecology* **81**, 8–14 (2000).

- Teng, J. & McCann, K. S. Dynamics of compartmented and reticulate food webs. *Am. Nat.* **164**, 86–100 (2004).
- Polis, G. A. & Strong, D. Food web complexity and community dynamics. *Am. Nat.* **147**, 813–846 (1996).
- Polis, G. A. & Holt, R. D. Intraguild predation: the dynamics of complex trophic interactions. *Trends Ecol. Evol.* **7**, 151–155 (1992).
- Paine, R. T. Food-web analysis through field measurement of per capita interaction strength. *Nature* **355**, 73–75 (1992).
- Wootton, J. T. Estimates and test of per capita interaction strength: diet, abundance, and impact of intertidally foraging birds. *Ecol. Monogr.* **67**, 45–64 (1997).
- Goldwasser, L. & Roughgarden, J. Construction and analysis of a large Caribbean food web. *Ecology* **74**, 1216–1233 (1993).
- Rafaeli, D. G. & Hall, S. J. in *Food Webs: Integration of Patterns & Dynamics* (eds Polis, G. A. & Winemiller, K. O.) 185–191 (Chapman & Hall, New York, 1996).
- Winemiller, K. O. Spatial and temporal variation in tropical fish trophic networks. *Ecol. Monogr.* **60**, 331–367 (1990).
- Polis, G. A. Complex trophic interactions in deserts: an empirical critique of food web theory. *Am. Nat.* **138**, 123–155 (1991).
- Moore, J. C. & Hunt, H. W. Resource compartmentation and the stability of real ecosystems. *Nature* **333**, 261–263 (1988).
- Moore, J. C., Walter, D. W. & Hunt, H. W. Habitat compartmentation and environmental correlates to food chain length. *Science* **243**, 238–240 (1989).
- Moore, J. C., de Ruiter, P. C. & Hunt, H. W. The influence of ecosystem productivity on food web stability. *Science* **261**, 906–908 (1993).
- Cohen, J. E. *Food Webs and Niche Space* (Princeton Univ. Press, Princeton, 1978).
- Ulanowicz, R. E. & Baird, D. Nutrient controls on ecosystem dynamics: the Chesapeake mesohaline community. *J. Mar. Syst.* **19**, 159–172 (1999).
- Baird, D. & Ulanowicz, R. E. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* **59**, 329–364 (1989).
- Neira, S. & Arancibia, H. Trophic interactions and community structure in the upwelling system off Central Chile (33–39 degrees S). *J. Exp. Mar. Biol. Ecol.* **312**, 349–366 (2004).
- Sanchez, F. & Olaso, I. Effects of fisheries on the Cantabrian Sea shelf ecosystem. *Ecol. Model.* **172**, 151–174 (2004).
- Trites, A. W., Livingston, P. A., Vasconcellos, M. C., Mackinson, S., Springer, A. M. & Pauly, D. Ecosystem change and the decline of marine mammals in the Eastern Bering Sea: testing the ecosystem shift and commercial whaling hypotheses. *Fisheries Centre Research Reports* **7**(1) (1999).
- Hunt, H. W. et al. The detrital food web in a shortgrass prairie. *Biol. Fertil. Soils* **3**, 57–68 (1987).
- de Ruiter, P. C., Neutel, A. & Moore, J. C. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* **269**, 1257–1260 (1995).
- Doles, J. A. *Survey of Soil Biota in the Arctic Tundra and their Role in Mediating Terrestrial Nutrient Cycling*. Masters thesis, Univ. of Northern Colorado (2000).
- Berg, M. P. *Decomposition, Nutrient Flow and Food Web Dynamics in a Stratified Pine Forest Soil*. Doctoral thesis, Vrije Universiteit, Amsterdam (1987).
- Vander Zanden, M. J. & Vadeboncoeur, Y. Fishes as integrators of benthic and pelagic food chains in lakes. *Ecology* **83**, 2152–2161 (2002).
- McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
- Tilman, D. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* **5**, 1455–1474 (1999).
- MacArthur, R. H. Fluctuations of animal populations and a measure of community stability. *Ecology* **36**, 533–536 (1955).
- McCann, K. S., Rasmussen, J. B. & Ulanowicz, R. E. The dynamics of spatially coupled food webs. *Ecol. Lett.* **8**, 513–523 (2005).
- Holt, R. D. Optimal foraging and the form of the predator isocline. *Am. Nat.* **122**, 521–541 (1983).
- Křivan, V. Optimal foraging and predator–prey dynamics. *Theor. Popul. Biol.* **49**, 265–290 (1996).
- Terborgh, J. et al. Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1925 (2001).
- Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. de Bruyn and J. Rasmussen for helpful comments. This work was supported by an NSERC grant to K.M., and by grants from the US National Science Foundation and the National Center for Ecological Analysis and Synthesis to J.C.M.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.R. (nrooney@uoguelph.ca).

ARTICLES

Mechanism of DNA translocation in a replicative hexameric helicase

Eric J. Enemark¹ & Leemor Joshua-Tor¹

The E1 protein of papillomavirus is a hexameric ring helicase belonging to the AAA+ family. The mechanism that couples the ATP cycle to DNA translocation has been unclear. Here we present the crystal structure of the E1 hexamer with single-stranded DNA discretely bound within the hexamer channel and nucleotides at the subunit interfaces. This structure demonstrates that only one strand of DNA passes through the hexamer channel and that the DNA-binding hairpins of each subunit form a spiral 'staircase' that sequentially tracks the oligonucleotide backbone. Consecutively grouped ATP, ADP and apo configurations correlate with the height of the hairpin, suggesting a straightforward DNA translocation mechanism. Each subunit sequentially progresses through ATP, ADP and apo states while the associated DNA-binding hairpin travels from the top staircase position to the bottom, escorting one nucleotide of single-stranded DNA through the channel. These events permute sequentially around the ring from one subunit to the next.

Replication initiation only occurs at specific DNA sites (origin of replication, *ori*), and can only occur once per cell division. Initiation begins with binding of *ori* by an origin-recognition protein such as bacterial DnaA (reviewed in ref. 1) or yeast ORC (reviewed in ref. 2). Subsequently, a helicase is assembled at *ori* and DNA polymerase(s), primase(s) and other factors are recruited to form the replisome, which processively synthesizes DNA complementarily to the template. In eukaryotes replication initiation involves many proteins. Even in *Escherichia coli*, at least three proteins (DnaA, DnaB and DnaC)¹ are required just to establish the replicative helicase at *ori*. Small DNA viruses such as papillomavirus, SV40 and AAV (adeno-associated virus) use a single initiator protein for origin recognition, helicase loading and helicase activity itself.

The viral initiator proteins E1, large T-antigen and Rep (from papillomavirus, SV40 and AAV, respectively) belong to helicase superfamily III (SF3; reviewed in ref. 3) and are members of the AAA+ (ATPases associated with cellular activities) family⁴. These helicases form hexameric rings that are believed to encircle substrate DNA and unwind it with 3' → 5' polarity^{5–7}. These properties resemble those of bacteriophage T7 gene product 4 (T7gp4) and *E. coli* DnaB, two hexameric replicative helicases that exhibit the opposite 5' → 3' polarity^{8,9}. Both T7gp4 and DnaB have been shown to encircle only one of the two DNA strands^{10,11}. The strand displacement assays used to determine the polarity of SF3 helicases indicates that the hexamers will load upon a single-stranded 3' tail^{7,12}, suggesting that the ring encircles this strand while sterically excluding the other. A mechanism where double-stranded DNA (dsDNA) passes through the channel after assembly at the *ori* was proposed for these helicases^{13,14}. However, the crystal structures of SV40 T-antigen (Tag)^{13,14} demonstrate channel diameters of <7 Å in the ATP-bound state and ~15 Å in the apo state in the AAA+ domains; this is insufficient to permit the passage of ~20-Å-diameter dsDNA, suggesting that only ssDNA can be accommodated.

Papillomavirus and SV40 have a closed circular genome, and consequently no ends are available for loading an intact ring. Whereas most DNA helicases require a region of ssDNA for entry, E1 and Tag can initiate unwinding from completely dsDNA, apparently by

causing helix melting and entry of the DNA helicase onto a single-stranded region. Sequence similarity between the carboxy-terminal halves of E1 and SV40 T-antigen, which contain the AAA+ domain, is significant. The amino-terminal halves of the proteins, which contain the DNA-binding domain (DBD), share little sequence similarity, but their structures are highly similar^{15,16}, reinforcing the likelihood that they use analogous mechanisms. The closed ring topology of their genomes suggests that the helicase is loaded as an open ring¹⁷, as proposed for DNA processivity clamps¹⁸, or as monomers sequentially until a hexameric ring encircling the DNA is produced^{17,19}. E1 carries out its transition from site-specific dsDNA binding (origin binding activity) to aspecific single-stranded binding and unwinding (helicase activity) via a transition of stoichiometry from a dimer to a double hexamer through a series of discrete intermediate oligomeric states assembled directly upon progressively distorted^{6,19–21} and ultimately melted origin DNA^{22,23}. To determine the topological and atomic details of DNA coordination by this group of ring helicases, and to elucidate features of the DNA translocation mechanism, we have determined the crystal structure of a bovine papillomavirus (BPV) E1 fragment comprising the AAA+ and oligomerization domains in complex with ssDNA, ADP and Mg²⁺.

Overall architecture

The crystal consists of two hexamers (subunits A–F for hexamer 1 and G–L for hexamer 2), each encircling a single strand of DNA (Fig. 1). The six oligomerization domains form a rigid collar with near proper six-fold rotational symmetry. The arrangement of the AAA+ domains resembles the arrangement in the Tag hexamer¹⁴ and in the predicted human papillomavirus type 18 (HPV-18) E1 AAA+ domain hexamer²⁴. However, unlike the oligomerization domains, in this case they deviate significantly from proper six-fold symmetry. Most of the protein–DNA interactions occur at hairpins located at the interior of the channel, which narrows to ~13 Å in diameter. These hairpins resemble a spiral staircase as they sequentially track the sugar-phosphate backbone of the ssDNA in a right-handed helical arrangement of a consistent set of interactions (Fig. 2). The

¹W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

subunits have varying modes of Mg^{2+} -ADP coordination, and each P-loop appears to have an associated ADP molecule except for subunit L.

DNA binding

All of the residues observed to interact with ssDNA are located at the interior of the hexameric ring on the AAA+ domains (Fig. 2). The K506 ammonium group interacts with one ssDNA phosphate oxygen while the H507 main-chain amide proton forms a hydrogen bond with the ssDNA phosphate of an adjacent nucleotide. The aliphatic portion of K506 and the aromatic groups of F464 and H507 form van der Waals interactions with the ssDNA sugar moiety linking these two phosphates. This set of interactions permutes sequentially around the hexameric ring with concurrent unitary ssDNA nucleotide progression between the subunits. The 5' end of the ssDNA is directed towards the N-terminal oligomerization domains, whereas the 3' end is directed towards the C terminus. The protein-DNA interactions are consistent between the two hexamers with one notable exception. One hexamer has five of the subunit hairpins engaged in ssDNA phosphate coordination whereas the other hexamer engages all six subunit hairpins. The details of this difference illustrate a logical transition within a cyclic translocation mechanism described below.

Intersubunit interactions and ADP binding

The most robust intersubunit interactions occur between the oligomerization domains through consistent interactions at each interface. The interactions between the AAA+ domains either mediate nucleotide binding or are involved in 'staircasing' the hairpins. As with Tag, the RFC (replication factor C) clamp loader, and other multisubunit AAA+ protein complexes, the amino acids responsible for nucleotide

coordination and hydrolysis are located on adjacent subunits. One contains the Walker A, Walker B and sensor-1 motifs. The adjacent subunit provides additional elements such as the arginine finger. For the E1-DNA hexamer, the intersubunit arrangement and separation varies. Consequently, the atomic details of nucleotide binding vary between the subunits (Fig. 3). In all cases, the Walker A and B motifs are consistently structured and generally coordinate the diphosphate group of an ADP molecule and an associated Mg^{2+} ion. In contrast, the adjacent subunit has three fundamental forms that correlate with the 'order' travelling around the ring as well as the vertical position of the DNA-binding hairpin in the staircase.

Form I is present at three of the interfaces in hexamer 1 and at two of the interfaces in hexamer 2 (see also Supplementary Fig. S1 and Supplementary Table S1). This form involves an intricate network of interactions between adjacent subunits (Fig. 3). Based on three factors, this form of nucleotide coordination is classified as ATP-type: (1) simultaneous engagement of all the AAA+ components implicated in ATP coordination and hydrolysis; (2) all of these elements structurally align with equivalents in the (Tag-ATP)₆ structure (Supplementary Fig. S2); and (3) a chloride ion sits at the anticipated position for an ATP γ -phosphate in three of the five cases (difference density is also observed at 2.5 σ and 3 σ in the other two cases but has not been modelled). The positions of K425 and R538 (the arginine finger) are major factors that define this nucleotide coordination mode. The ammonium group of K425 concurrently binds the α - and β -phosphate in a position very similar to the sensor-2 arginine, R217, of RuvB, an AAA+ protein involved in Holliday branch migration. As a result, we classify this lysine as 'sensor-2'. The arginine finger, R538, simultaneously binds a Walker B aspartate of the first subunit, D479, and a water molecule that coordinates the Mg^{2+} ion. Notably, the guanidinium group of R538

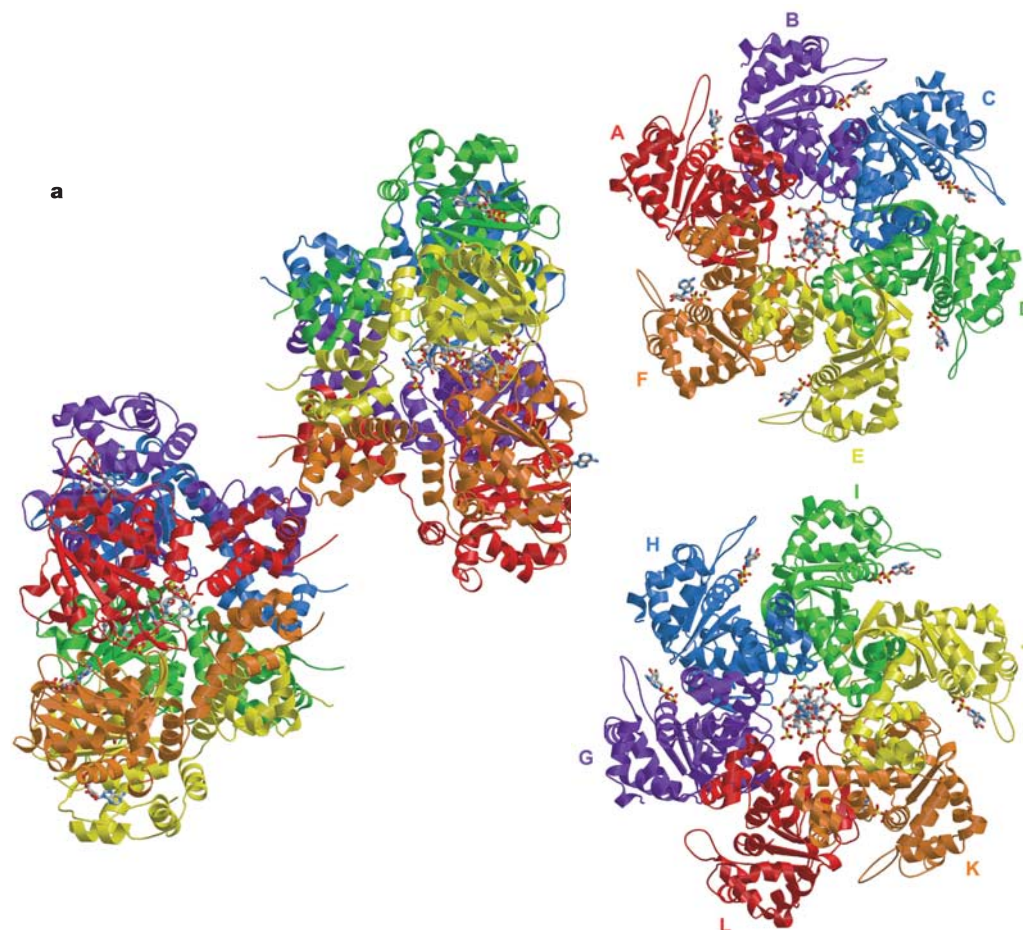


Figure 1 | Structure of the E1 hexameric helicase in complex with ssDNA and ADP. **a, b,** Ribbon representations of the two E1 helicase hexamers viewed perpendicular (**a**) and parallel (**b**) to the channels. Individual subunits are colour-coded with hexamer 1 on top and hexamer 2 on the bottom. A crystal lattice interaction between the hexamers is illustrated (**a**). The oligomerization domains form a rigid collar, located between the AAA+ domains in **a** and projected towards the reader in **b**. The DNA in the central channel and the ADP molecules at subunit interfaces are depicted in stick representation. Images were prepared with Bobscript³⁴ and Raster3D^{35,36}.

is appropriately positioned to interact with the γ -phosphate of an ATP, generally occupied by a Cl^- ion in the crystal structure. The Cl^- ion also interacts with sensor-1 of the first subunit, N523. Several residues on a connected pair of helices participate in this coordination. Arginine R493, denoted here as 'sensor-3', interacts with D489 as well as the Walker B aspartate, D479, and sensor-1, N523. Tyrosine Y499 interacts with the ADP α - PO_4 .

Form II uses the Walker A and B motifs of one subunit, but only Y499 of the adjacent subunit (Fig. 3). In this form, most of the ATP-type interactions are absent and display flexibility. R493 interacts with Walker B D479 and sensor 1 N523 of the neighbouring subunit for the interface between subunits D and E of hexamer 1 and the I–J interface of hexamer 2. However, it is too distant for the E–F (hexamer 1) and J–K (hexamer 2) interfaces (see also Supplementary Fig. S1). A striking difference between this form and the ATP type described above is the sandwiching of Y534 between R493 and R538 (arginine finger). This interaction places R538 too distant to interact with a γ - PO_4 of an ATP molecule. The greatly reduced interactions and complete exclusion of the arginine finger led us to classify this as 'ADP-type' coordination. No Cl^- ions are observed in this form.

Form III does not use any residues of the adjacent subunit and also binds the Mg^{2+} ion differently than the preceding forms (Fig. 3). This form has been classified as 'apo' due to the total lack of involvement of the adjacent subunit. In the apo-type coordination, an ADP molecule is usually present at the Walker A and B motif regions, but it is coordinated differently than above. In the K–L interface, the α - PO_4 and β - PO_4 of the ADP molecule adopt a reverse orientation relative to the other subunit interfaces.

Correlation of nucleotide state with hairpin height

The DNA-binding hairpins form a right-handed staircase that sequentially tracks the ssDNA backbone. The vertical position on the staircase for the hairpin of a given subunit correlates with the type of nucleotide coordination. The subunits that participate in ATP-type coordinations place their hairpins at the top of the staircase; the hairpins of apo-type subunits occupy the bottom of the staircase. The hairpins of ADP-type subunits lie at the intermediate positions. The staircase is maintained by a consistent set of interactions (Fig. 2) between the hairpins of adjacent subunits involving K506, which also contacts a ssDNA phosphate as described above. The ammonium group of this lysine interacts with three sites on the adjacent hairpin: the acidic group of D504 and the main-chain carbonyl groups of R505 and K508. These staircasing elements are conserved in all E1 sequences as well as in Tag and AAV, and we expect these proteins to exhibit the same sets of interactions and form a staircase of hairpins when ssDNA and multiple nucleotide states are present simultaneously (Supplementary Fig. S3). Previous crystal structures of Tag, which lacked DNA and possessed a single nucleotide state^{13,14}, did not display a staircase of hairpins. Interestingly, a hexameric ring structure of T7gp4, which had multiple nucleotide states but lacked DNA²⁵, also demonstrated a staircase of the DNA-binding hairpins via a lysine (K467, chain C) and an acidic residue (D470, chain B). These interactions were maintained between subunits with assigned ATP-type coordination. Also, these staircasing elements appear to be conserved in DnaB, as shown from a sequence alignment with T7gp4 (Supplementary Fig. S4). The adoption of a staircase by DNA-binding hairpins through a set of interactions involving a

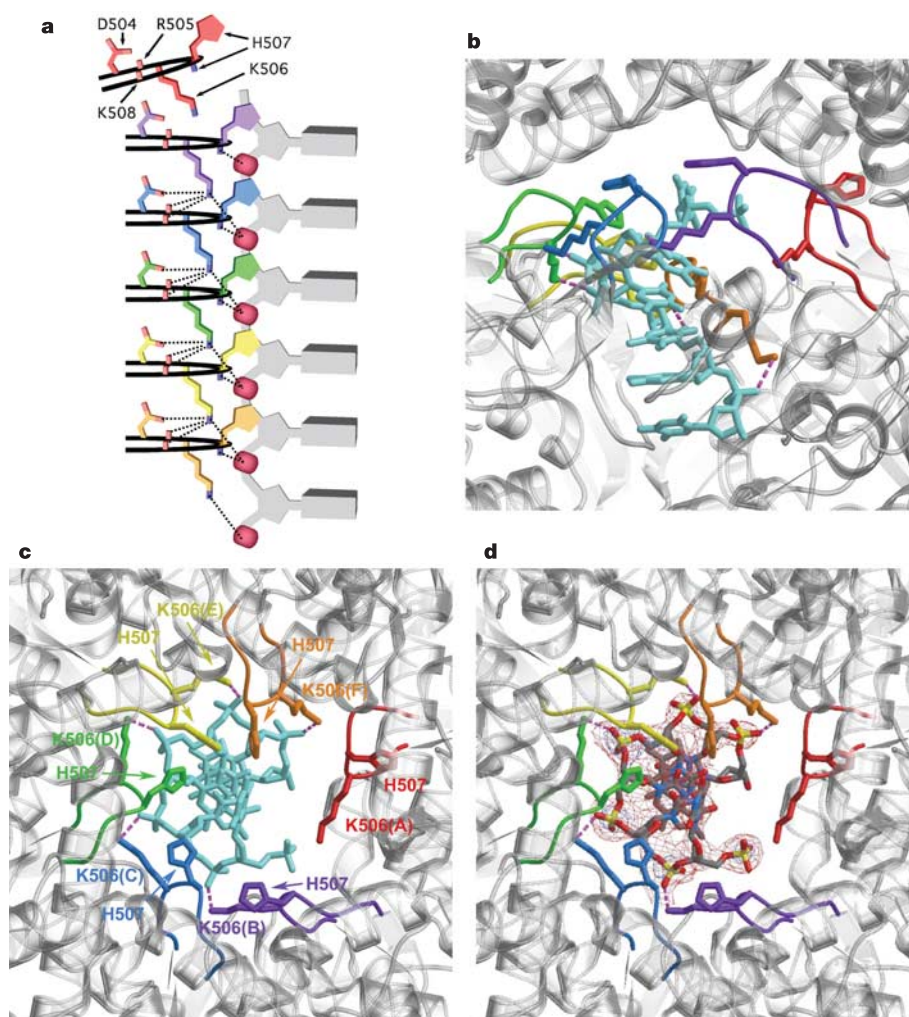


Figure 2 | DNA binding in the E1 hexameric helicase.

a–c, Interactions involving the AAA+ hairpins and ssDNA shown as a schematic (**a**) and as a ribbon diagram in two perpendicular views (**b**, **c**) for hexamer 1. K506 coordinates a ssDNA phosphate, and the main-chain amide of H507 interacts with the phosphate of an adjacent nucleotide. K506 also mediates interactions with three sites on the adjacent hairpin to form the staircase of hairpins—the side chain of D504 and the main chain carbonyl groups of R505 and K508 (the latter two side chains are omitted for clarity). Hydrogen bonds are drawn as dotted (**a**) or dashed (**b–d**) lines. Hairpins of the individual subunits are colour-coded as in Fig. 1. In **b** and **c** only H507 and K506 side chains are shown. The letters in parentheses identify the subunits. Other components of the protein are coloured grey with transparency. The DNA is in light blue. **d**, Same as **c** with superimposed $F_o - F_c$ difference electron density calculated before inclusion of any DNA in the model. The electron density is contoured at 3σ (red) and at 6σ (blue).

basic and an acidic residue of a neighbouring subunit may be a fundamental feature of hexameric helicases.

The correlation of nucleotide state with vertical position of the DNA-binding hairpins is consistent with previous structural observations for Tag in the absence of DNA^{13,14}. Apart from the presence of DNA, the major difference here is that these states are all present in the same hexamer.

The crystal structure of hexameric T7gp4 bound to the non-hydrolysable ATP analogue ADPNP in the absence of DNA also

exhibits multiple nucleotide-binding configurations with an associated trend for the DNA-binding loops²⁵. A crystallographic two-fold-related set of three sites is observed, two with associated ADPNP molecules (ATP-type and ADP-type) and one empty.

A coordinated escort mechanism for DNA translocation

The hexameric structures determined here display varying nucleotide coordination modes that correlate with the vertical position of the associated DNA-binding hairpins. The ssDNA nucleotides are

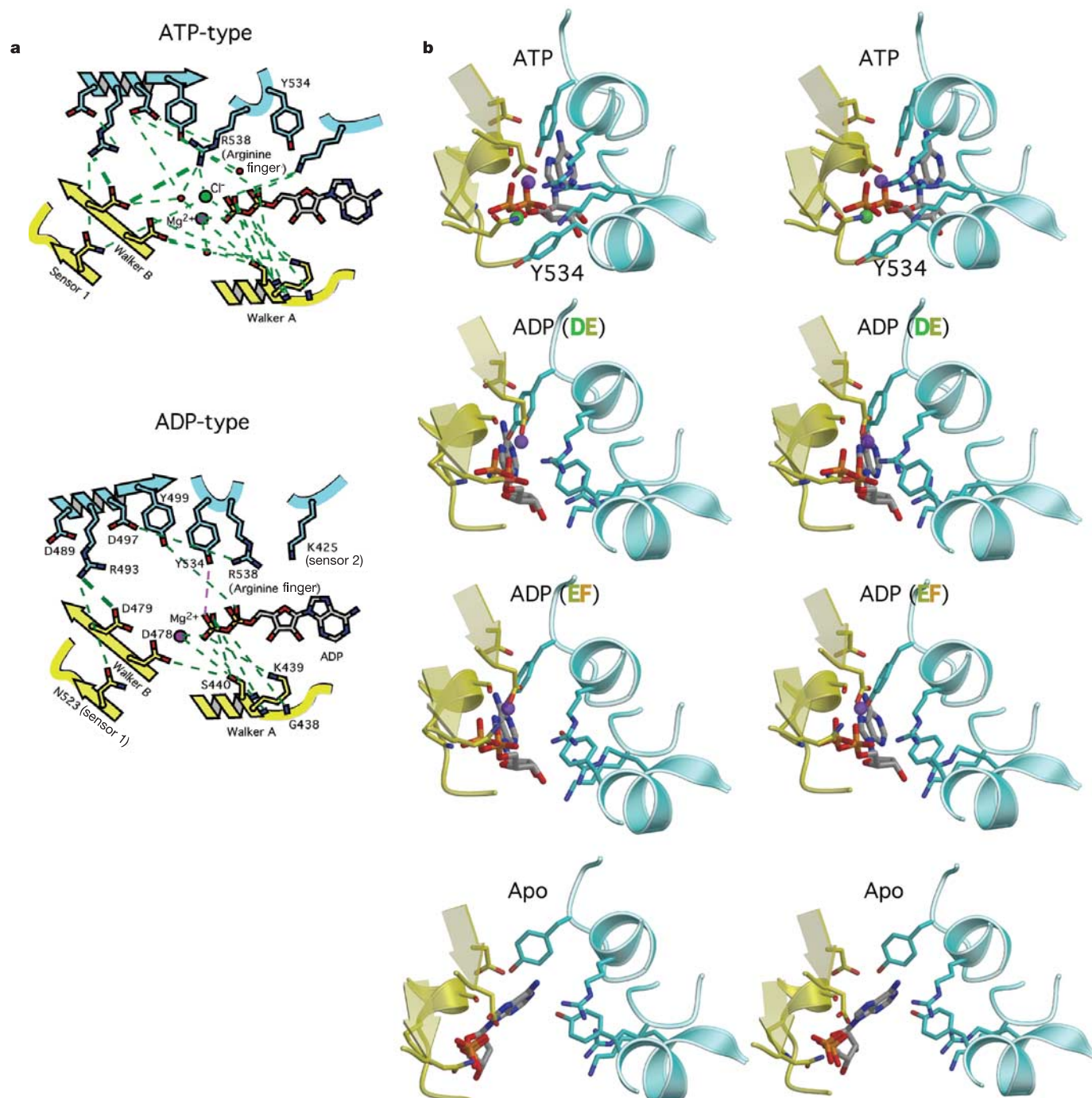


Figure 3 | Intersubunit interactions and nucleotide binding. **a, b.** The hydrogen-bonding networks in the intersubunit nucleotide coordination types shown as a schematic (**a**) and in stereoview (**b**). The subunit containing the Walker A, Walker B and sensor-1 motifs is coloured yellow; the adjacent subunit containing the arginine finger is coloured cyan. Mg²⁺ ions are coloured purple and Cl⁻ ions are green. The ATP-like coordination (top) has multiple interactions and engages R538 and K425. A Cl⁻ ion is

generally observed at the anticipated position for a γ -PO₄. The ADP-like configuration has fewer interactions, and the position of Y534 between R494 and R538 precludes the interactions observed in the ATP-like case. K425 is too distant to interact with the nucleotide. Successive ADP-type configurations (middle) show increasing separation between the subunits. The apo configuration is shown at the bottom.

arranged with a 1 nucleotide per subunit increment. As a result, a straightforward and compelling DNA translocation mechanism integrating these details can be inferred (Fig. 4). Each DNA-binding hairpin maintains continuous contact with one unique nucleotide of ssDNA and migrates downward via ATP hydrolysis and subsequent ADP release at the subunit interfaces. ATP hydrolysis occurs between subunits located towards the top of the staircase, while ADP release occurs between subunits located towards the bottom. The downward movements of the hairpins are coordinated by the staircasing interactions described above. The hairpin at the bottom of the staircase releases its associated ssDNA phosphate to conclude its journey through the hexameric channel. Upon binding a new ATP molecule, this subunit moves to the top of the staircase to pick up the next available ssDNA phosphate, initiating its escorted journey through the channel and repeating the process. For one full cycle of the hexamer, each subunit hydrolyses 1 ATP molecule, releases 1 ADP molecule and translocates 1 nucleotide through the interior channel. A full cycle, therefore, translocates 6 nucleotides with associated hydrolysis of 6 ATPs and release of 6 ADPs.

Both hexamers display a noticeable gap between the AAA+ domains located at the top and bottom positions of the staircase (Fig. 1b). This interface is where the most sizeable movements occur in the mechanism proposed here. Upon ATP binding, the subunit at the bottom of the staircase moves to the top position and closes this gap. Simultaneously, a new gap opens between this subunit and the previous subunit (now bottom). The two hexamers in the crystal display different configurations at this interface, as demonstrated by the different number of ATP-like (and empty) states and the number of hairpins engaged in ssDNA coordination (Supplementary Table S1). For hexamer 1, an ATP-type configuration is observed between subunits A and B, but the hairpin of subunit A is not engaged in ssDNA coordination. Subunit A is consistent with a subunit that has recently bound an ATP molecule, but has not yet bound DNA (see also Fig. 2b, c). This state is not present in hexamer 2, where the subunit at the top of the staircase has an ATP state and binds to ssDNA. Hexamer 2, however, displays two empty nucleotide states whereas hexamer 1 has only one. Subunit L is the lowest on the hexamer 2 staircase and has no counterpart in hexamer 1. It represents the state that directly precedes ATP binding. This is the only subunit in the structure where no ADP is observed at the P-loop, and it displays much larger thermal parameters in the AAA+ domain than the other subunits.

This mechanism has some similarities to the previously proposed operation of T7gp4. For the previous T7gp4 proposal, successive loops also bind to successive nucleotides but pass them through the channel by handing them off from one loop to the next in a “bucket

brigade” manner²⁶. In such a mechanism, the loops would maintain a nearly fixed height as the individual nucleotides are passed downward from one loop to the next. In contrast, in the escort mechanism proposed here, each hairpin maintains a continuous set of interactions with one nucleotide, and the entire unit collectively migrates downward. We suggest that the T7gp4 hairpins also co-migrate with a given nucleotide in an analogous coordinated escort mechanism. The assigned nucleotide configurations in the crystal structure of T7gp4 in complex with ADPNP²⁵ place the DNA-binding loops of the ATP-assigned coordination farthest from the primase domains, while the empty state places them closest, in agreement with the probable translocation of DNA towards the primase domains.

For T7gp4, subunits engaged in ATP coordination display a higher affinity for DNA than the ADP and empty states²⁷. The same relative order of affinities in the E1 hexamer would lead to an additional driving force for the coordinated escort mechanism of cyclic DNA translocation because subunits at the top of the staircase would have a higher affinity for DNA (more prone to bind) than those at the bottom of the staircase (more prone to release). Binding of a new ATP molecule by a subunit in the empty state at the bottom of the staircase would serve not only to return it to the top of the staircase, but also to increase its affinity for DNA.

Translocation direction

The mechanism described here directly couples ATP hydrolysis to DNA translocation by using ATP hydrolysis to drive the sequential movement of the hairpins and pump ssDNA in that direction. This mechanism is consistent with the demonstrated 3' → 5' polarity for SF3 helicases and the observed polarity of the DNA in the structure with the 5' end located towards the top (oligomerization domain side) of the complex. Previously, the opposite translocation direction was proposed for Tag¹³, where the same correlation of hairpin height with nucleotide configuration was identified in the absence of DNA. DNA was proposed to translocate from the empty associated state (bottom) towards the ATP-associated state (top). A drawback of that mechanism is that it does not derive any work from ATP hydrolysis. Such a mechanism is energetically driven by ATP binding. ATP hydrolysis would only serve to facilitate return to the empty state.

Oligomerization domain as processivity factor

The rigid collar formed by the oligomerization domains could serve as a single-stranded equivalent of double-stranded processivity factors such as polymerase sliding clamps like PCNA. Whereas PCNA holds dsDNA topologically together, the static ring formed by six oligomerization domains in the E1 helicase keeps two strands topologically apart. This function would be important to prevent the ring from falling off the substrate DNA because the AAA+ domains have relatively weak interactions between the subunits, including a significant gap between the subunits with the bottom and top hairpins. The robust intersubunit interactions of the ring formed by the oligomerization domains ensure that the hexamer continues to surround the given single strand.

Final remarks

A translocation mechanism in which ATP is hydrolysed sequentially from one subunit to the next with the staircase migrating downward may constitute a general mechanism for hexameric helicases. Although the three nucleotide-binding configurations observed in the structure define relative positions for the ATP, ADP and empty states in a coordinated escort mechanism, the precise details regarding the sequence and timing of ATP hydrolysis, phosphate release, ADP release and ATP binding in this mechanism remain to be determined. This mechanism might not operate by one exclusive timing of these events; slight variations might operate in parallel as described for T7gp4 (ref. 28). The localization of the double-stranded/single-stranded fork junction in this complex also remains

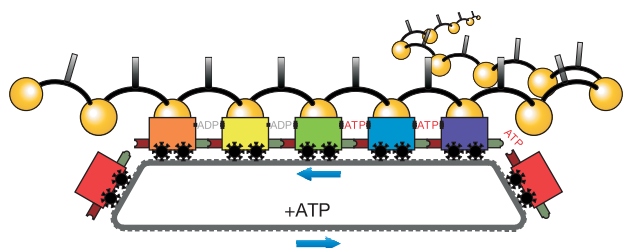


Figure 4 | Cartoon depiction of a coordinated escort mechanism for the E1 hexameric helicase. Each subunit is depicted as a wagon (or boxcar), colour-coded as in Fig. 1. Each wagon transports one DNA nucleotide from the right side to the left. Staircasing interactions are depicted by the wagon couplers. ATP hydrolysis and ADP release occur along the path as depicted by the nucleotide coordination type between the wagons. The red, leftmost wagon ejects its associated DNA nucleotide and returns to the right side upon binding a new ATP molecule. This wagon then picks up the next DNA nucleotide of the series, couples to the purple wagon, and carries its cargo towards the left. Figure prepared by J. Duffy.

unclear. This species may involve the DBD, which in this model is anticipated to be proximal to the dsDNA entering the complex.

The assembly of the helicase at the *ori* relies upon the DBD. The ultimate helicase species established at the *ori* is a double hexamer²⁹. Our structure suggests that the double hexamer consists of two individual hexamers that encircle opposing single strands. The helicase activity of this species results from unidirectional translocation along the encircled strand with steric exclusion of the other, as described for other helicases such as the RNA helicase NPH-II³⁰. The mechanism described here can be used to establish two different species where the hexamers remain associated or move apart, as seen by electron microscopy^{31,32} (see Supplementary Fig. S5). Further studies are necessary to address these issues.

METHODS

Complex preparation. A fragment encoding BPV-1 E1 residues 306–577 was generated by polymerase chain reaction (PCR) amplification from an expression plasmid template provided by A. Stenlund, and cloned into pGEX-4T-1 (Amersham). The GST fusion protein was expressed in *E. coli* and purified in a manner analogous to methods described previously for other E1 GST fusion proteins¹⁵, and was concentrated in the presence of 5 mM ADP and 200 mM MgCl₂.

Crystallization and data collection. The sample was mixed with excess oligo-dT oligonucleotide (13-mer), and immediately used for crystallization trials. Crystals grew at 17 °C by the hanging-drop method with a well solution consisting of 100 mM KCl. Crystals were flash frozen in a stream of cold nitrogen gas (100 K) at beamline X26C of the NSLS for evaluation of diffraction. The most favourable sample was transferred to liquid nitrogen and moved to beamline X29 for data collection. Data collection statistics are shown in Supplementary Table S2.

Structure determination and refinement. The structure was solved by molecular replacement with the program PHASER³³ by placing 12 copies of a hybrid search model consisting of the previously reported helicase domain of HPV-18 E1 (ref. 24) (Protein Data Bank code 1TUE) and the helicase domain of BPV E1 (E.J.E., T. J. Takara, A. Stenlund and L.J., unpublished structure of E1(368–577) refined to 1.65 Å). Further details on the data collection, structure determination and refinement can be found in Supplementary Information.

Received 8 April; accepted 1 June 2006.

- Messer, W. The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* **26**, 355–374 (2002).
- Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
- Hickman, A. B. & Dyda, F. Binding and unwinding: SF3 viral helicases. *Curr. Opin. Struct. Biol.* **15**, 77–85 (2005).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Seo, Y. S., Muller, F., Lusky, M. & Hurwitz, J. Bovine papillomavirus (BPV) encoded E1 protein contains multiple activities required for BPV DNA replication. *Proc. Natl Acad. Sci. USA* **90**, 702–706 (1993).
- Yang, L. *et al.* The E1 protein of the papillomavirus BPV-1 is an ATP dependent DNA helicase. *Proc. Natl Acad. Sci. USA* **90**, 5086–5090 (1993).
- Sedman, J. & Stenlund, A. The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities. *J. Virol.* **72**, 6893–6897 (1998).
- Ahnert, P. & Patel, S. S. Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. *J. Biol. Chem.* **272**, 32267–32273 (1997).
- LeBowitz, J. H. & McMacken, R. The *Escherichia coli* dnaB replication protein is a DNA helicase. *J. Biol. Chem.* **261**, 4738–4748 (1986).
- Egelman, E. H., Yu, X., Wild, R., Hingorani, M. M. & Patel, S. S. Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proc. Natl Acad. Sci. USA* **92**, 3869–3873 (1995).
- Kaplan, D. L. & O'Donnell, M. DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands. *Mol. Cell* **10**, 647–657 (2002).
- Seo, Y. S. & Hurwitz, J. Isolation of helicase alpha, a DNA helicase from HeLa cells stimulated by a fork structure and signal-stranded DNA-binding proteins. *J. Biol. Chem.* **268**, 10282–10295 (1993).
- Gai, D., Zhao, R., Li, D., Finkelstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47–60 (2004).
- Li, D. *et al.* Structure of the replicative helicase of the oncoprotein SV40 large tumor antigen. *Nature* **423**, 512–518 (2003).
- Enemark, E. J., Chen, G., Vaughn, D. E., Stenlund, A. & Joshua-Tor, L. Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus. *Mol. Cell* **6**, 149–158 (2000).
- Luo, X., Sanford, D. G., Bullock, P. A. & Bachovchin, W. W. Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nature Struct. Biol.* **3**, 1034–1039 (1996).
- Davey, M. J. & O'Donnell, M. Replicative helicase loaders: ring breakers and ring makers. *Curr. Biol.* **13**, R594–R596 (2003).
- Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. Clamp loaders and sliding clamps. *Curr. Opin. Struct. Biol.* **12**, 217–224 (2002).
- Enemark, E. J., Stenlund, A. & Joshua-Tor, L. Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex. *EMBO J.* **21**, 1487–1496 (2002).
- Gillette, T. G., Lusky, M. & Borowiec, J. A. Induction of structural changes in the bovine papillomavirus type 1 origin of replication by the viral E1 and E2 proteins. *Proc. Natl Acad. Sci. USA* **91**, 8846–8850 (1994).
- Sanders, C. M. & Stenlund, A. Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor. *EMBO J.* **17**, 7044–7055 (1998).
- Schuck, S. & Stenlund, A. Assembly of a double hexameric helicase. *Mol. Cell* **20**, 377–389 (2005).
- Chen, G. & Stenlund, A. Sequential and ordered assembly of E1 initiator complexes on the papillomavirus origin of DNA replication generates progressive structural changes related to melting. *Mol. Cell. Biol.* **22**, 7712–7720 (2002).
- Abbate, E. A., Berger, J. M. & Botchan, M. R. The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. *Genes Dev.* **18**, 1981–1996 (2004).
- Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).
- Crampton, D. J., Mukherjee, S. & Richardson, C. C. DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol. Cell* **21**, 165–174 (2006).
- Hingorani, M. M. & Patel, S. S. Interactions of bacteriophage T7 primase/helicase protein with single-stranded and double-stranded DNAs. *Biochemistry* **32**, 12478–12487 (1993).
- Liao, J. C., Jeong, Y. J., Kim, D. E., Patel, S. S. & Oster, G. Mechanochemistry of T7 DNA helicase. *J. Mol. Biol.* **350**, 452–475 (2005).
- Fouts, E. T., Yu, X., Egelman, E. H. & Botchan, M. R. Biochemical and electron microscopic image analysis of the hexameric E1 helicase. *J. Biol. Chem.* **274**, 4447–4458 (1999).
- Kawaoka, J., Jankowsky, E. & Pyle, A. M. Backbone tracking by the SF2 helicase NPH-II. *Nature Struct. Biol.* **11**, 526–530 (2004).
- Wessel, R., Schweizer, J. & Stahl, H. Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *J. Virol.* **66**, 804–815 (1992).
- Lin, B. Y., Makhov, A. M., Griffith, J. D., Broker, T. R. & Chow, L. T. Chaperone proteins abrogate inhibition of the human papillomavirus (HPV) E1 replicative helicase by the HPV E2 protein. *Mol. Cell. Biol.* **22**, 6592–6604 (2002).
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
- Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134 (1997).
- Bacon, D. J. & Anderson, W. F. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220 (1988).
- Merritt, E. A. & Murphy, M. E. P. Raster3D version 2.0 - A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Robinson (beamline X29) and A. Héroux (beamline X26C) for support with data collection at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. We also thank G. Hannon and A. Gann for critical reading of the manuscript, and B. Stillman, N. Tolia and members of the Joshua-Tor laboratory for discussions. The NSLS is supported by the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences. This work was supported by an NIH grant to L.J.

Author Information Coordinates and structure factors are deposited in the Protein Data Bank under accession code 2GXA. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.J. (leemor@cshl.edu).

LETTERS

An X-ray-emitting blast wave from the recurrent nova RS Ophiuchi

J. L. Sokoloski¹, G. J. M. Luna^{1,2}, K. Mukai³ & Scott J. Kenyon¹

Stellar explosions such as novae and supernovae produce most of the heavy elements in the Universe. The onset of a nova is well understood¹ as driven by runaway thermonuclear fusion reactions on the surface of a white dwarf in a binary star system; but the structure, dynamics and mass of the ejecta are not well known. In rare cases, the white dwarf is embedded in the wind nebula of a red-giant companion, and the explosion products plough through the nebula and produce X-ray emission. Here we report X-ray observations of such an event, from the eruption of the recurrent nova RS Ophiuchi^{2,3}. The hard X-ray emission from RS Ophiuchi early in the eruption emanates from behind a blast wave, or outward-moving shock wave, that expanded freely for less than 2 days and then decelerated owing to interaction with the nebula. The X-rays faded rapidly, suggesting that the blast wave deviates from the standard spherical shell structure^{4–6}. The early onset of deceleration indicates that the ejected shell had a low mass, the white dwarf has a high mass⁷, and that RS Ophiuchi is therefore a progenitor of the type of supernova (type Ia) integral to studies of the expansion of the Universe.

On 12 February 2006, RS Ophiuchi (RS Oph) was discovered in outburst near the historical maximum optical brightness of 4.4 mag (ref. 8)—more than four times brighter than the threshold for naked-eye visibility. The Rossi X-Ray Timing Explorer (RXTE) satellite⁹ first detected strong hard X-rays from RS Oph—the strongest ever observed from a white dwarf—at peak X-ray brightness on outburst day 2 with the All Sky Monitor camera. RXTE subsequently measured the X-ray spectrum with the Proportional Counter Array (PCA) on outburst days 3, 6, 10, 14, 17 and 21 (Fig. 1). During this time, the binary was in the phase of its orbit in which the white dwarf and red giant appear side-by-side from the perspective of the Earth^{10,11}.

During the first three weeks of the outburst, the total X-ray luminosity from the hottest X-ray-emitting gas fell from $\sim 300L_{\odot}$ to $\sim 30L_{\odot}$ (integrated from 0.5 to 20 keV, and taking a distance of 1.6 kpc (refs 12, 13); $L_{\odot}$ is the solar luminosity. From day 6 onward, the X-ray emission from the hottest gas decreased as $t^{-5/3}$, where t is the time since the beginning of the outburst (Fig. 2a). After rising to maximum in about 2 days, the emission measure (n^2V , where n is the density of radiating electrons and V is the X-ray-emitting volume) remained roughly constant at $\sim 10^{58} \text{ cm}^{-3}$ until day 6, after which time it fell as $t^{-4/3}$. For comparison, the X-ray luminosity from hot gas in classical novae, where the white dwarf is not embedded in a dense nebula, is always below $30L_{\odot}$ (refs 14, 15), and the classical nova with the most comprehensive X-ray observations¹⁶ (V1974 Cyg) took more than 100 days to reach a maximum X-ray emission measure of a few times 10^{56} cm^{-3} . As the outburst progressed and the X-ray luminosity decreased, the bulk of the X-ray emission from RS Oph shifted to lower energies (Fig. 1). This shift corresponds to a

drop in the post-shock temperature. From the time of the first spectral observation on day 3, this temperature decreased as $t^{-2/3}$ (Fig. 2b).

If the X-rays are from circumstellar material heated by the blast wave produced in the nova explosion, we can associate the characteristic X-ray temperature with the blast-wave speed via the standard strong-shock relation. Optical emission-line widths directly measure the motion of either gas behind the blast-wave shock that has undergone charge exchange¹⁷ or the ejecta, and these line widths¹⁸ support the shock speeds that we derive. As $u \propto T^{1/2}$, where T is the post-shock plasma temperature and u is the blast-wave speed, the rate of temperature decrease indicates that the blast-wave speed decreased as $t^{-1/3}$. This direct measurement of the shock deceleration rate is, to our knowledge, the first for any nova. For an initial shock speed of $3,500 \text{ km s}^{-1}$ (refs 19, 20), and extrapolating backwards from our inferred shock speeds, we conclude that the deceleration began at approximately day 1.7. With the measured deceleration, we can obtain the blast-wave radius as a function of time. Comparing the blast-wave radius on days 21 and 27 to radio images of the expanding shock on those days (M. P. Rupen, A. J. Mioduszewski and J.L.S., manuscript in preparation), we confirm the 1.6 kpc distance to RS Oph.

The standard blast-wave theory presupposes that the shock structure is self-similar, or that the density and temperature as a function of fractional distance from the white dwarf to the shock front are similar to these functions at later times. In the initial ejecta-dominated phase, the ejecta expand freely and produce a shock moving at constant velocity. After the expanding blast wave has swept up a few times the ejecta mass, it begins to decelerate and enters the Sedov-Taylor phase. After the 1985 outburst of RS Oph, standard supernova blast-wave theory was adapted to explain the late X-ray emission⁶. Models for RS Oph must also account for the fact that the nova explosion takes place within a stellar wind whose density far from the binary decreases as the inverse of the distance from the binary squared ($n \propto 1/r^2$, where r is the distance from the wind-producing red giant). Whereas the early hard X-ray emission from classical novae is thought to arise in the ejecta that has been heated by a reverse shock²¹, the early hard X-ray emission from RS Oph comes predominantly from circumstellar material heated by the forward-moving blast wave.

Our observations indicate that the standard self-similar blast-wave model does not explain the ejection of material from this nova. Although our discovery that $u \propto t^{-1/3}$ agrees with the theoretical prediction for a blast wave moving into a stellar wind and the general expansion law for a shock wave resulting from a point explosion^{4–6}, the observed rate of X-ray fading disagrees with predictions. In the current outburst, the initial X-ray flux was almost 100 times brighter than expected on the basis of observations and modelling of the

¹Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ²Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão 1226, Cid. Universitária 05508-900, São Paulo, Brazil. ³NASA Goddard Space Flight Center and Universities Space Research Association, Code 662, Greenbelt, Maryland 20771, USA.

late-time X-ray emission in the 1985 outburst^{6,22}. The total X-ray flux for thermal bremsstrahlung emission from a fully ionized gas is proportional to $T^{1/2}n^2V$, where T is temperature. If the volume of the post-shock emitting region increases as the radius of the blast wave cubed (as in the standard similarity solutions for supernova shells), the X-rays should decrease as t^{-1} . The X-ray emission from RS Oph in the Sedov-Taylor phase decayed instead as $t^{-5/3}$.

The X-ray flux evolution reflects the environment of the binary and suggests how the nova explosion deviates from a spherically symmetric, self-similar blast wave. The fast X-ray rise of RS Oph compared to classical novae is due to the higher circumstellar density and the faster initial shock speed. The observed peak emission measure can be attained in 2 days if the material swept up by the blast wave produces thermal bremsstrahlung emission at the immediate post-shock density and temperature. The observed X-ray decay rate from day 6 onward is consistent with emission from a hot post-shock region whose volume increases roughly as the square of the blast-wave radius. Such an emitting-volume evolution could be produced by radiative cooling, which has little impact on the early-time shock velocity, but a significant impact on the post-

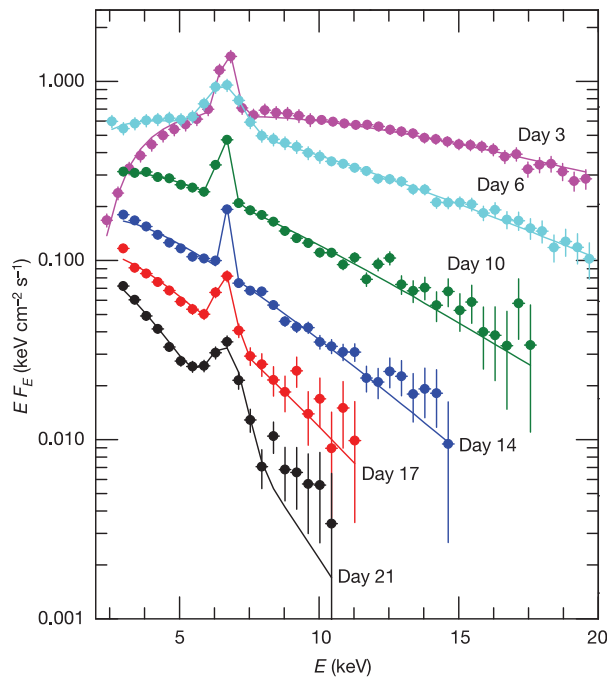


Figure 1 | X-ray spectra from the first 3 weeks of the 2006 outburst of RS Oph. These six spectra were taken with the PCA instrument on board the RXTE satellite. The abscissa is the energy, E , of the detected X-rays. The quantity EF_E is the photon energy times the energy flux density, and it provides a measure of the relative amount of energy being emitted across the spectrum. The usable energy range for the PCA is approximately 2–25 keV. The spectra were reduced using standard data reduction software (the FTOOLS package) and the standard RXTE background models. The presence of a blended emission line near 6.6 keV from H-like and He-like Fe indicates that the X-ray emission is optically thin thermal plasma emission; the spectra can all be reasonably well fitted (reduced χ^2 ranging from 0.4 to 1.7 for the six observations) with a single-temperature thermal bremsstrahlung model plus line emission from Fe and absorption by intervening material. Although RXTE is not very sensitive to absorption, which preferentially affects photons with energies less than around 2 keV, the absorption was high enough on day 3 ($N_H = (5.5 \pm 1.1) \times 10^{22} \text{ cm}^{-2}$, where N_H is the column density of neutral hydrogen) to have a significant impact on the low-energy end of the spectrum. The absorption lessened in the subsequent observations as the blast wave moved outward and the amount of neutral intervening wind decreased. After day 10, we fixed the absorption parameter in our fits to the interstellar value^{13,22,30} of $N_H = 4 \times 10^{21} \text{ cm}^{-2}$. Error bars represent 1 standard deviation.

shock temperature structure²³. Alternatively, or in conjunction with radiative cooling, a non-spherical ejecta geometry such as that inferred from radio observations^{24,25}, or a contribution to the X-ray luminosity from ejecta heated by a reverse shock, could also steepen the X-ray decline.

Finally, because the transition to the Sedov-Taylor phase occurs when a few times the ejecta mass have been swept up, the timing of the onset of deceleration constrains the ejected shell mass independently of all previous estimates. Taking a density inside the binary of 10^9 cm^{-3} from observational constraints on the red-giant mass-loss rate¹⁰ and assuming that the density fall-off does not begin until outside the binary, on about day 2 the shock had swept up only a few times $10^{-7} M_\odot$ of material ($M_\odot$ is the solar mass). Thus, the ejecta mass could not have been much more than $10^{-7} M_\odot$. Theoretical models of nova explosions with recurrence times of 20 years, as in RS Oph, have been constructed for white dwarfs with masses of $1.25 M_\odot$ and $1.40 M_\odot$ (ref. 7). Whereas the $1.25 M_\odot$ models require an ejecta mass of $\sim 10^{-6} M_\odot$, the $1.40 M_\odot$ models eject material with a total mass of $2 \times 10^{-7} M_\odot$. The ejected shell mass of $\sim 10^{-7} M_\odot$ from RS Oph is thus more consistent with the $1.4 M_\odot$ model.

The mass accumulation efficiency of white dwarfs in recurrent novae is controversial^{7,26}. Our conclusion that the white dwarf in RS Oph must be extremely close to the maximum mass shows that significant mass accumulation is possible, and that some recurrent novae will explode as type Ia supernovae. If recurrent novae lead to type Ia supernovae such as the one recently discovered to contain hydrogen in its spectrum²⁷, sweeping up of the red-giant wind by the nova blast wave could explain the cavity surrounding this supernova^{28,29}. Our results also have bearing on studies of type II supernovae. The early days of the 2006 outburst of RS Oph show the first two phases of blast-wave evolution, the analogue of which can take hundreds of years for a supernova remnant. These data also provide the blast-wave deceleration law, which is difficult to measure directly in supernovae. Our results should motivate modelling that will lead to better understanding of cooling mechanisms as

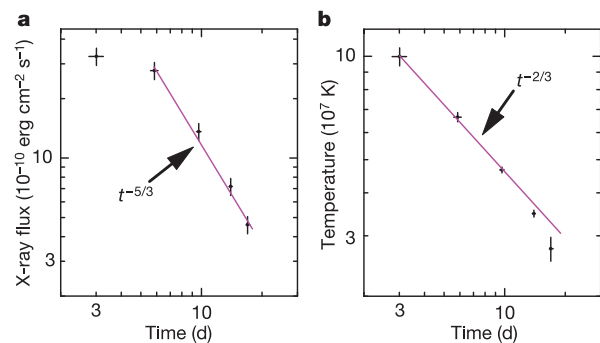


Figure 2 | X-ray flux and post-shock plasma temperature as a function of time. We measure the time, t , from the start of the outburst, which we take to be February 12.6 UT. **a**, X-ray flux in energy range 0.5–20 keV from the hottest plasma, obtained by integrating the best-fit thermal bremsstrahlung model, corrected for the effects of absorption by intervening material. This energy range was chosen to match typical observing ranges and to include most of the emission from the hot plasma. The uncertainty was taken to be 5%, and comes primarily from uncertainty in the model parameters, especially interstellar absorption. **b**, The characteristic X-ray plasma temperature determined from fitting the X-ray spectra with thermal bremsstrahlung models, as a function of time. Error bars represent 1 standard deviation. The temperature decays as expected for material that has been shock-heated by a decelerating Sedov-Taylor blast wave moving into a $n \propto 1/r^2$ stellar wind. As RXTE is sensitive to X-rays with energies greater than 2 keV, we obtain the best determination of the parameters from the first several observations, when the post-shock plasma is hottest. By the sixth observation, the error bars have become too large for these data to be useful (Fig. 1), so the data from the final observation are not included on the plots.

well as post-shock and ejecta structure of novae and other stellar explosions.

Received 21 March; accepted 11 May 2006.

1. Starrfield, S., Truran, J. W. & Sparks, W. M. CNO abundances and hydrodynamic studies of the Nova outburst. V–1.0-solar-mass models with small mass envelopes. *Astrophys. J.* **226**, 186–202 (1978).
2. Kenyon, S. J. *The Symbiotic Stars* (Cambridge University Press, Cambridge, UK, 1986).
3. Bode, M. F. (ed.) *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (VNU Science, Utrecht, 1987).
4. Sedov, L. *Similarity and Dimensional Methods in Mechanics* (Academic, New York, 1959).
5. Chevalier, R. A. Self-similar solutions for the interaction of stellar ejecta with an external medium. *Astrophys. J.* **258**, 790–797 (1982).
6. Bode, M. F. & Kahn, F. D. A model for the outburst of nova RS Ophiuchi in 1985. *Mon. Not. R. Astron. Soc.* **217**, 205–215 (1985).
7. Yaron, O., Prialnik, D., Shara, M. M. & Kovetz, A. An extended grid of nova models. II. The parameter space of nova outbursts. *Astrophys. J.* **623**, 398–410 (2005).
8. Narumi, H. et al. RS Ophiuchi. *IAU Circ.* **8671** (2006).
9. Jahoda, K. et al. In-orbit performance and calibration of the Rossi X-ray Timing Explorer (RXTE) Proportional Counter Array (PCA). *Proc. SPIE* **2808**, 59–70 (1996).
10. Dobrzycka, D. & Kenyon, S. J. A new spectroscopic orbit for RS Ophiuchi. *Astron. J.* **108**, 2259–2266 (1994).
11. Fekel, F. C., Joyce, R. R., Hinkle, K. H. & Skrutskie, M. F. Infrared spectroscopy of symbiotic stars. I. Orbits for well-known S-type systems. *Astron. J.* **119**, 1375–1388 (2000).
12. Hjellming, R. M. et al. Radio observations of the 1985 outburst of RS Ophiuchi. *Astrophys. J.* **305**, L71–L75 (1986).
13. Snijders, M. A. J. in *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (ed. Bode, M. F.) 51–60 (VNU Science, Utrecht, 1987).
14. Mukai, K. & Ishida, M. The early X-ray emission from V382 Velorum (Nova Velorum 1999): An internal shock model. *Astrophys. J.* **551**, 1024–1030 (2001).
15. Greiner, J., Orio, M. & Schartel, N. XMM-Newton observations of Nova LMC 2000. *Astron. Astrophys.* **405**, 703–710 (2003).
16. Orio, M. in *Compact Binaries in the Galaxy and Beyond* 182–186 (eds Tovmassian, G. & Sion, E.) (IAU Colloquium 194, Instituto de Astronomia, Universidad Nacional Autonoma de Mexico, Ciudad Universitaria, 2004).
17. Chevalier, R. A., Kirshner, R. P. & Raymond, J. C. The optical emission from a fast shock wave with application to supernova remnants. *Astrophys. J.* **235**, 186–195 (1980).
18. Iijima, T. RS Ophiuchi. *IAU Circ.* **8675** (2006).
19. Buil, C. RS Ophiuchi. *Cent. Bur. Astron. Tel.* **403** (2006).
20. Fujii, M. Optical spectrum of RS Oph. *VSNET-alert* **8869** (2006).
21. O'Brien, T. J., Lloyd, H. M. & Bode, M. F. An interacting winds model for the X-ray emission from V838 Her (Nova Herculis 1991). *Mon. Not. R. Astron. Soc.* **271**, 155–160 (1994).
22. Mason, K. O., Córdova, F. A., Bode, M. F. & Barr, P. in *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (ed. Bode, M. F.) 167–175 (VNU Science, Utrecht, 1987).
23. O'Brien, T. J. & Kahn, F. D. Models for the remnants of recurrent novae—II. Dynamical effect of radiative heat loss. *Mon. Not. R. Astron. Soc.* **228**, 277–287 (1987).
24. Taylor, A. R., Davis, R. J., Porcas, R. W. & Bode, M. F. VLBI observations of RS Oph—a recurrent nova with non-spherical ejection. *Mon. Not. R. Astron. Soc.* **237**, 81–91 (1989).
25. Lloyd, H. M., Bode, M. F., O'Brien, T. J. & Kahn, F. D. Bipolar recurrent nova outbursts—I. Hydrodynamic models. *Mon. Not. R. Astron. Soc.* **265**, 457–470 (1993).
26. Hachisu, I. & Kato, M. Recurrent novae as a progenitor system of Type Ia supernovae. I. RS Ophiuchi subclass: Systems with a red giant companion. *Astrophys. J.* **558**, 323–350 (2001).
27. Hamuy, M. et al. An asymptotic-giant-branch star in the progenitor system of a type Ia supernova. *Nature* **424**, 651–654 (2003).
28. Wood-Vasey, W. M., Wang, L. & Aldering, G. Photometry of SN 2002ic and implications for the progenitor mass-loss history. *Astrophys. J.* **616**, 339–345 (2004).
29. Wood-Vasey, W. M. & Sokoloski, J. L. Novae as a mechanism for producing cavities around the progenitors of SN 2002ic and other SNe Ia. *Astrophys. J.* (in the press).
30. Davis, R. J. in *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (ed. Bode, M. F.) 187–192 (VNU Science, Utrecht, 1987).

Acknowledgements We are grateful for discussions with J. Raymond, C. Rakowski, M. Wood-Vasey and C. Matzner. J.L.S. acknowledges support from the NSF. G.J.M.L. acknowledges support from NASA, and from CNPq and FAPESP (Brazil).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.L.S. (jsokoloski@cfa.harvard.edu).

An asymmetric shock wave in the 2006 outburst of the recurrent nova RS Ophiuchi

T. J. O'Brien¹, M. F. Bode², R. W. Porcas³, T. W. B. Muxlow¹, S. P. S. Eyres⁴, R. J. Beswick¹, S. T. Garrington¹, R. J. Davis¹ & A. Evans⁵

Nova outbursts¹ take place in binary star systems comprising a white dwarf and either a low-mass Sun-like star or, as in the case of the recurrent nova RS Ophiuchi², a red giant. Although the cause of these outbursts is known to be thermonuclear explosion of matter transferred from the companion onto the surface of the white dwarf³, models of the previous (1985) outburst of RS Ophiuchi failed to adequately fit the X-ray evolution⁴ and there was controversy over a single-epoch high-resolution radio image, which suggested that the remnant was bipolar^{5,6} rather than spherical as modelled. Here we report the detection of spatially resolved structure in RS Ophiuchi from two weeks after its 12 February 2006 outburst. We track an expanding shock wave as it sweeps through the red giant wind, producing a remnant similar to that of a type II supernova but evolving over months rather than millennia⁷. As in supernova remnants, the radio emission is non-thermal (synchrotron emission), but asymmetries and multiple emission components clearly demonstrate that contrary to the assumptions of spherical symmetry in models of the 1985 explosion, the ejection is jet-like, collimated by the central binary whose orientation on the sky can be determined from these observations.

During the previous outburst of RS Ophiuchi (RS Oph) in 1985 a campaign was organized incorporating observations ranging from radio to X-ray wavelengths. The results included the detection of bright, evolving X-ray emission from hot gas suggested to arise from the expanding shock wave⁸. This time we have monitored RS Oph from much earlier in the outburst, both in X-rays^{9–13}, and at radio wavelengths with the Multi-Element Radio-Linked Interferometer Network (MERLIN), the Very Large Array (VLA), the Very Long Baseline Array (VLBA) and the European VLBI Network (EVN)^{14–16}.

Here we concentrate on the early Very Long Baseline Interferometry (VLBI) and MERLIN imaging observations that resolve the expanding radio source (Fig. 1). In the first epoch (13.8 d after outburst, taking day 0 as 2006 February 12.83; ref. 17) the radio emission takes the form of an approximately circular structure, which is significantly brighter on its eastern side. This one-sided ring initially expands at a speed of about 0.62 mas d^{-1} (Fig. 2), equivalent to $1,730 \text{ km s}^{-1}$ in the plane of the sky at our assumed distance of 1,600 pc. However, the structure quickly grows more complex with the appearance of a second component to the east of the ring. Subsequent 5-cm MERLIN imaging also shows the appearance of a third component to the west (Fig. 3). In 1985, RS Oph was followed at radio wavelengths from 18 d to a year after outburst^{18,19}. Only one attempt was made (77 d after outburst) to obtain a VLBI image^{5,6}, interpreted as a three-component radio source extending east–west to $\sim 200 \text{ mas}$. As only three telescopes of the early EVN had

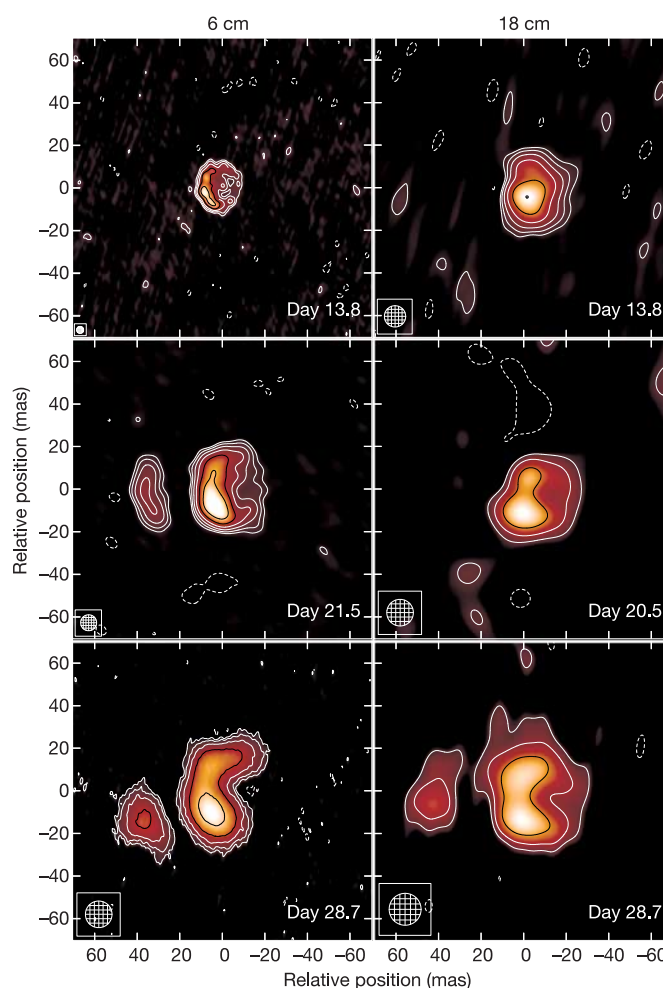


Figure 1 | High-resolution radio images of RS Oph. Images of RS Oph at wavelengths of 6 cm (left column) and 18 cm (right column) made with the VLBA on 2006 February 26 (day 13.8) and March 13 (day 28.7), and the EVN on March 5/6 (day 20.5/21.5). In each case, north is up and east is to the left, and the images are restored with the circular beams shown at lower left. The first 6-cm VLBA image has a resolution of 3.3 mas (5 AU at a distance of 1,600 pc) and a peak flux density of $4.7 \text{ mJy beam}^{-1}$ corresponding to a brightness temperature of around $4 \times 10^7 \text{ K}$. The contour levels are $(-1, 1, 2, 4, 8, 16, 32)$ times a base level given by 227, 105 and $220 \mu\text{Jy beam}^{-1}$ for the 6-cm images (in increasing time order), and 500, 900 and $900 \mu\text{Jy beam}^{-1}$ for the 18-cm images.

¹Jodrell Bank Observatory, School of Physics & Astronomy, The University of Manchester, Macclesfield, Cheshire, SK11 9DL, UK. ²Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Birkenhead, CH41 1LD, UK. ³Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany. ⁴Centre for Astrophysics, University of Central Lancashire, Preston, PR1 2HE, UK. ⁵Astrophysics Group, School of Physical & Geographical Sciences, Keele University, Staffordshire, ST5 5BG, UK.

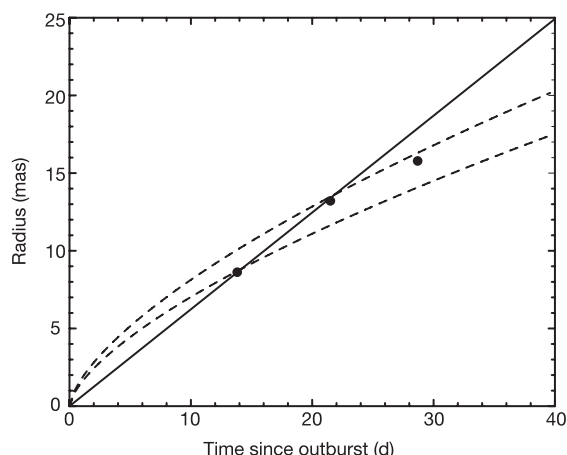


Figure 2 | Expansion of the bright ring-like component. The plotted points are estimates of the radius of the ring-like component from the 6-cm images. The solid line drawn through the origin and the first data point corresponds to free expansion at a speed of 0.62 mas d^{-1} , equivalent to $1,730 \text{ km s}^{-1}$ at our assumed distance of 1,600 pc (ref. 19). However, existing models⁸ suggest that after about day 6 the shock should be decelerating as it sweeps up red giant wind with its radius given by $at^{2/3}$ (here a is a constant, and t is time since outburst). Comparison with the X-ray emission seen in 1985⁴ led to estimates for a of $2.03 \times 10^{10} \text{ cm s}^{-2/3}$ and $2.35 \times 10^{10} \text{ cm s}^{-2/3}$, shown as the dashed lines. These are broadly consistent with the observations. However, although the expansion is clear, the measurements are sufficiently difficult that at this stage we are not able to clearly distinguish free expansion from the predicted deceleration. The 6-cm image from 26 February is clearly ring-like and a radius can easily be measured, but in the following two epochs the analysis is harder as the ring is only one-sided. Here we adopt the following method. In each 6-cm image, a series of north–south slices show two peaks, one from each side of the ring. The plotted points are half the separation of these peaks, determined by fitting gaussian profiles to the slice showing the largest separation. Formal errors on these measurements are typically around 0.2 mas, but these underestimate the true uncertainty. For example, analysis of the expanding shell of supernova 1993J²⁹ shows that the peaks may not represent the true location of the shock wave as they will be a convolution of the underlying emission with the beam.

been used, it was unclear whether this reflected the true nature of the emission. However, our results suggest that the source is developing into a similar structure during the current outburst.

Our measurement of the expansion speed of the ring is consistent with velocity estimates from the Doppler-broadened line profiles seen at optical²⁰, infrared²¹ and X-ray¹⁰ wavelengths and in ultraviolet spectroscopy²² of the 1985 outburst. Assuming this corresponds to the velocity of the shock wave resulting from the explosion, then gas would be heated to $\sim 4 \times 10^7 \text{ K}$, consistent with temperatures in the range $(0.2\text{--}7) \times 10^7 \text{ K}$ inferred from X-ray observations with the Swift, Chandra and RXTE satellites^{10,13,23}. This scenario is also consistent with our observed brightness temperatures. Taking the ratio of mass-loss rate (into the red giant wind) to wind velocity to be $6 \times 10^{12} \text{ g cm}^{-1}$ (from models of the 1985 outburst⁴), we estimate the density in the shell 14 d after outburst to be around $10^{-17} \text{ g cm}^{-3}$. A shell with these densities and a temperature of around $10^7\text{--}10^8 \text{ K}$ has a very small optical depth to free–free radiation. In such a case, the observed brightness temperatures are much lower than the electron temperature, and so thermal emission from hot gas alone is insufficient to explain the observed level of radiation. We therefore conclude that there is a significant contribution from a non-thermal mechanism, probably synchrotron radiation from particles accelerated in the shock wave in a manner similar to that inferred for supernova remnants.

The one-sided appearance of the ring is likely to be due to the density distribution in the red giant wind and/or asymmetries in the explosion. Notably, the second component is visible at first only in

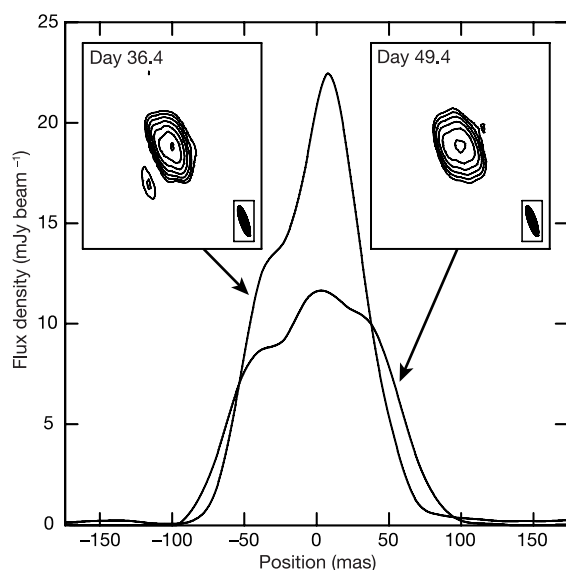


Figure 3 | A third component seen in MERLIN radio imaging. East–west slices through the MERLIN maps (at a wavelength of 5 cm, shown as insets) of RS Oph from 21 March (day 36.4) and 3 April (day 49.4). These show that on day 36.4 the source comprised a bright component (which we identify with the main ring in the images of Fig. 1) and a fainter component to the east (the feature which first appears in the EVN map of day 21.5). By day 49.4 the slice shows that MERLIN has detected a third component to the west. The contours in the maps are at $(-1, 1, 2, 4, 8, 16, 32, 64) \times 326 \mu\text{Jy beam}^{-1}$, and the beam size is $121 \times 31 \text{ mas}$.

the 6-cm image but a week later is also visible at 18 cm, indicating its spectrum has flattened, consistent with a source that is initially subject to free–free absorption (greater at longer wavelengths) in some overlying material. The likely source of such absorption is the ionized red giant wind. Figure 4 presents a simple model, where the nova explosion results in a bipolar shock-heated shell that expands through the red giant wind. The bipolarity is the result of either an inherent asymmetry in the mass ejection from the white dwarf (that is, a jet-like outflow), or an equatorial enhancement in the red giant wind that leads to shaping of the shock wave in a manner similar to the interacting stellar winds model for planetary nebulae²⁴. Whatever the origin of the bipolarity, it is reasonable to expect its symmetry axis to be perpendicular to the plane of the orbit of the central binary system. The inclination of the binary orbit to the line of sight is estimated to be $30\text{--}40^\circ$ from radial velocity observations before the 2006 outburst²⁵, so we adopt a value of 35° . Figure 4 shows model images at around the time of the VLBI imaging and the later MERLIN imaging. The early one-sided structure is quite clear, and the model predicts that as the source expands, the free–free absorption reduces, so the structure becomes more symmetrical with a central bright source between two other components, just as the MERLIN imaging (Fig. 3) is suggesting and as was seen at day 77 in 1985^{5,6}.

Although the binary inclination will remain the same from one outburst to the next, the observed structure may also depend on the relative orientation of the binary components at the time of outburst. Orbital elements determined from infrared spectroscopy²⁶ suggest that the 2006 outburst took place when the red giant and white dwarf appear most widely separated. With an orbital period of 456 d this is about the same phase as the 1985 outburst 16.9 orbits earlier. Hence we expect the geometrical properties to be comparable, consistent with the similar radio structures seen in the two outbursts.

The model also explains the rapid rise in the radio emission—detected by MERLIN only 4.4 d after outburst¹⁴, and projecting back to a turn-on at day 3.8—as the result of rapidly reducing absorption in the central, densest, regions of the red giant wind. However, a

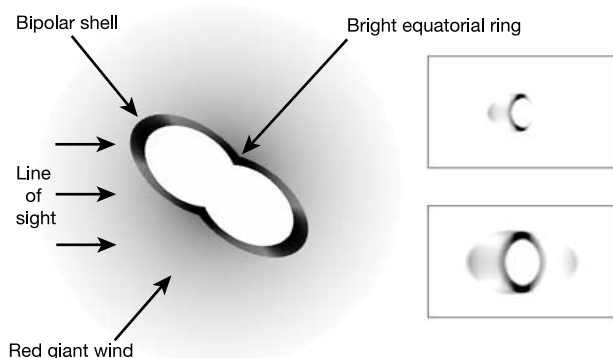


Figure 4 | A bipolar model for the radio emission. Main figure, a slice showing gas density in a simple model for the structure of RS Oph. A bipolar shell of shock-heated gas, swept up by the ejecta, is embedded within a sphere filled with 21 yr of mass-loss from the red giant. The gas density in the red giant wind, given by the 1985 model fits⁴, is proportional to r^{-2} (where r is radius) and its temperature is 10^4 K. The density in the shocked shell is simply assumed to be four times the wind density at that radius, but with some enhancement towards the poles, and its temperature is 5×10^7 K. The gas emits via a combination of free-free and flat-spectrum synchrotron emission in which we assume equipartition between the energy in relativistic electrons (a fixed small fraction of the thermal energy) and the magnetic field. The emission is integrated along a line of sight given by the inclination of the binary system allowing for intrinsic free-free absorption, mostly provided by the ionized red giant wind. Synthetic images are shown as insets, oriented for comparison with the observations. The central ring is the waist of the shell, which is the densest, and hence brightest, part. In the first epoch (top inset), the furthest half of the ring actually suffers less absorption owing to the swept-out cavity along this line of sight, and so appears brighter. The second component, first seen at day 21.5, is the nearer lobe of the shell with the farther lobe initially obscured. At the second epoch (bottom inset) the shell has expanded, the absorption is reduced and the appearance has become more symmetrical. We identify the third component seen in the second MERLIN map of Fig. 3 with the unveiling of the more distant lobe.

uniformly expanding shell is not consistent with the slow movement of the second component. More detailed analysis, including a physically self-consistent description of the density and temperature distribution within the shell and recognition that the explosion on the white dwarf does not lie at the centre of the red giant wind, requires a full hydrodynamical simulation, which we leave for a paper to follow.

These observations provide an elegant confirmation of the basic shock model for RS Ophiuchi developed after the 1985 outburst, but provide new challenges in understanding the unexpected asymmetries seen in the radio shell. Their development into a two-sided structure is the best evidence thus far for the existence of jets in a recurrent nova. The mechanisms needed to produce jets in a wide variety of sources (ranging from active galactic nuclei to young stellar objects) are still unclear, but the favoured model suggests production at the centre of accretion disks with hydromagnetic acceleration and collimation²⁷. This is consistent with our model, in which the symmetry axis of the bipolar shell is orthogonal to the plane of the binary orbit.

Further epochs of high-resolution radio imaging, combined with ongoing X-ray monitoring, will enable the shock dynamics in a nova explosion to be studied in unprecedented detail, including modelling the departure from spherical symmetry. As well as providing additional insights into important physical processes such as particle acceleration, this will lead to improved estimates of the ejected mass and outburst energy, essential to constraining the thermonuclear

runaway model and to determining whether RS Oph and objects of its type will eventually explode as type Ia supernovae²⁸.

Received 6 April; accepted 31 May 2006.

- Bode, M. F. & Evans, A. (eds) *Classical Novae* (Wiley, Chichester, 1989).
- Bode, M. F. (ed.) *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (VNU Science, Utrecht, 1987).
- Starrfield, S., Truran, J. W. & Sparks, W. M. The nova outburst: thermonuclear runaways on degenerate dwarfs. *New Astron. Rev.* **44**, 81–86 (2000).
- O'Brien, T. J., Bode, M. F. & Kahn, F. D. Models for the remnants of recurrent novae III—Comparison with the X-ray observations of RS Ophiuchi (1985). *Mon. Not. R. Astron. Soc.* **255**, 683–693 (1992).
- Porcas, R. W., Davis, R. J. & Graham, D. A. in *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (ed. Bode, M. F.) 203–207 (VNU Science, Utrecht, 1987).
- Taylor, A. R., Davis, R. J., Porcas, R. W. & Bode, M. F. VLBI observations of RS Oph – a recurrent nova with non-spherical ejection. *Mon. Not. R. Astron. Soc.* **237**, 81–91 (1989).
- Bode, M. F. & Kahn, F. D. A model for the outburst of nova RS Ophiuchi in 1985. *Mon. Not. R. Astron. Soc.* **217**, 205–215 (1985).
- Mason, K. O., Cordova, F. A., Bode, M. F. & Barr, P. in *RS Ophiuchi (1985) and the Recurrent Nova Phenomenon* (ed. Bode, M. F.) 167–175 (VNU Science, Utrecht, 1987).
- Bode, M. F. *et al.* RS Ophiuchi. *IAU Circ.* **8675** (2006).
- Ness, J.-U. *et al.* RS Ophiuchi. *IAU Circ.* **8683** (2006).
- Osborne, J. *et al.* The decline and rise of the X-ray flux of RS Oph. *Astron. Tel.* **764** (2006).
- Osborne, J. *et al.* RS Oph: Swift X-ray observations find short period modulation and highly variable low energy flux. *Astron. Tel.* **770** (2006).
- Bode, M. F. *et al.* Swift observations of the 2006 outburst of the recurrent nova RS Ophiuchi I. Early X-ray emission from the shocked ejecta and red giant wind. *Astrophys. J.* (submitted); preprint at (<http://arXiv.org/astro-ph/0604618>) (2006).
- Eyres, S. P. S., O'Brien, T. J., Muxlow, T. W. B., Bode, M. F. & Evans, A. RS Ophiuchi. *IAU Circ.* **8678** (2006).
- O'Brien, T. J. *et al.* RS Ophiuchi. *IAU Circ.* **8684** (2006).
- O'Brien, T. J. *et al.* RS Ophiuchi. *IAU Circ.* **8688** (2006).
- Hirosawa, K., Narumi, H. & Kanai, K. RS Ophiuchi. *Cent. Bur. Electron. Tel.* **399** (2006).
- Padin, S., Davis, R. J. & Bode, M. F. RS Ophiuchi – First radio detection of a recurrent nova outburst. *Nature* **315**, 306 (1985).
- Hjellming, R. M. *et al.* Radio observations of the 1985 outburst of RS Ophiuchi. *Astrophys. J.* **305**, L71–L75 (1986).
- Buil, C. RS Ophiuchi. *Cent. Bur. Electron. Tel.* **403** (2006).
- Evans, A. *et al.* RS Ophiuchi. *IAU Circ.* **8682** (2006).
- Shore, S. N., Kenyon, S. J., Starrfield, S. & Sonneborn, G. On the interpretation of the ultraviolet spectra of symbiotic stars and recurrent novae II. The 1985 outburst of RS Ophiuchi. *Astrophys. J.* **456**, 717–737 (1996).
- Sokoloski, J. L., Luna, G. J. M. & Mukai, K. Fourth RXTE observation of RS Oph in outburst: continued fading. *Astron. Tel.* **754** (2006).
- Lloyd, H. M., Bode, M. F., O'Brien, T. J. & Kahn, F. D. Bipolar recurrent nova outbursts I. Hydrodynamic models. *Mon. Not. R. Astron. Soc.* **265**, 457–470 (1993).
- Dobrzycka, D. & Kenyon, S. J. A new spectroscopic orbit for RS Ophiuchi. *Astron. J.* **108**, 2259–2266 (1994).
- Fekel, F. C., Joyce, R. R., Hinkle, K. H. & Skrutskie, M. F. Infrared spectroscopy of symbiotic stars I. Orbits for well-known S-type systems. *Astron. J.* **119**, 1375–1388 (2000).
- Livio, M., Pringle, J. E. & King, A. R. The disk-jet connection in microquasars and active galactic nuclei. *Astrophys. J.* **593**, 184–188 (2003).
- Sokoloski, J. L., Luna, G. J. M., Mukai, K. & Kenyon, S. J. An X-ray-emitting blast wave from the recurrent nova RS Ophiuchi. *Nature* doi:10.1038/nature04893 (this issue).
- Marcaide, J. M. *et al.* Deceleration in the expansion of SN 1993J. *Astrophys. J.* **486**, L31–L34 (1997).

Acknowledgements We are grateful to the staff of the VLBA (a facility of the NSF operated under cooperative agreement by Associated Universities, Inc.), EVN (a joint facility of European, Chinese, South African and other radio astronomy institutes funded by their national research councils), the NASA DSN facility at Robledo and MERLIN (a National Facility of PPARC operated by the University of Manchester) for help in obtaining these observations at short notice.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J.O.B. (tim.obrien@manchester.ac.uk).

LETTERS

Graphene-based composite materials

Sasha Stankovich^{1*}, Dmitriy A. Dikin^{1*}, Geoffrey H. B. Dommett¹, Kevin M. Kohlhaas¹, Eric J. Zimney¹, Eric A. Stach³, Richard D. Piner¹, SonBinh T. Nguyen² & Rodney S. Ruoff¹

Graphene sheets—one-atom-thick two-dimensional layers of sp^2 -bonded carbon—are predicted to have a range of unusual properties. Their thermal conductivity and mechanical stiffness may rival the remarkable in-plane values for graphite ($\sim 3,000 \text{ W m}^{-1} \text{ K}^{-1}$ and $1,060 \text{ GPa}$, respectively); their fracture strength should be comparable to that of carbon nanotubes for similar types of defects^{1–3}; and recent studies have shown that individual graphene sheets have extraordinary electronic transport properties^{4–8}. One possible route to harnessing these properties for applications would be to incorporate graphene sheets in a composite material. The manufacturing of such composites requires not only that graphene sheets be produced on a sufficient scale but that they also be incorporated, and homogeneously distributed, into various matrices. Graphite, inexpensive and available in large quantity, unfortunately does not readily exfoliate to yield individual graphene sheets. Here we present a general approach for the preparation of graphene-polymer composites via complete exfoliation of graphite⁹ and molecular-level dispersion of individual, chemically modified graphene sheets within polymer hosts. A polystyrene-graphene composite formed by this route exhibits a percolation threshold¹⁰ of ~ 0.1 volume per cent for room-temperature electrical conductivity, the lowest reported value for any carbon-based composite except for those involving carbon nanotubes¹¹; at only 1 volume per cent, this composite has a conductivity of $\sim 0.1 \text{ S m}^{-1}$, sufficient for many electrical applications¹². Our bottom-up chemical approach of tuning the graphene sheet properties provides a path to a broad new class of graphene-based materials and their use in a variety of applications.

Graphite oxide is a layered material produced by the oxidation of graphite (Fig. 1a). In contrast to pristine graphite, the graphene-derived sheets in graphite oxide (graphene oxide sheets) are heavily oxygenated, bearing hydroxyl and epoxide functional groups on their basal planes, in addition to carbonyl and carboxyl groups located at the sheet edges^{13,14}. The presence of these functional groups makes graphene oxide sheets strongly hydrophilic, which allows graphite oxide to readily swell and disperse in water^{2,15,16}. Previous studies by our team have shown that a mild ultrasonic treatment of graphite oxide in water results in its exfoliation to form stable aqueous dispersions that consist almost entirely of 1-nm-thick sheets (Fig. 1b)⁹. Given the uniformity of the observed thicknesses and that platelets one-half of the observed minimum thickness are never detected (or any other inverse integer value, such as one-third, and so on), we believe that these represent fully exfoliated graphene oxide sheets. In fact, at present, exfoliation of graphite oxide is the only way to produce stable suspensions of quasi-two-dimensional carbon sheets, making this a strategic starting point for large-scale synthesis of graphene sheets. As such, graphite oxide has recently attracted attention as a filler for polymer nanocomposites^{17–19}.

Nanocomposites have been heralded²⁰ as a 'radical alternative to

conventional filled polymers or polymer blends,' especially in the fields of transportation and electronics. Unfortunately, owing to their hydrophilic nature, graphene oxide sheets can only be dispersed in aqueous media that are incompatible with most organic polymers. In addition, graphite oxide is electrically insulating, unlike graphite, which limits its usefulness for the synthesis of conductive nanocomposites. It has been demonstrated^{19,18,21,22}, however, that the electrical conductivity of graphite oxide can be significantly increased by chemical reduction, presumably owing to the restoration of a graphitic network of sp^2 bonds. But reduction of exfoliated graphene oxide nanoplatelets in water results in their irreversible coagulation⁹, which then makes dispersion within a polymer matrix at the individual sheet level impossible.

We recently demonstrated that the exfoliation behaviour of graphite oxide can be altered by changing the surface properties of graphene oxide sheets by way of chemical functionalization. A number of chemically modified graphite oxides were prepared by treating graphite oxide with organic isocyanates²³. The isocyanate treatment reduces the hydrophilic character of graphene oxide sheets by forming amide and carbamate ester bonds to the carboxyl and hydroxyl groups of graphite oxide, respectively. As a result, such isocyanate-derivatized graphite oxides no longer exfoliate in water but readily form stable dispersions in polar aprotic solvents (such as *N,N*-dimethylformamide (DMF)), consisting of completely exfoliated, functionalized individual graphene oxide sheets with thickness $\sim 1 \text{ nm}$, as determined by atomic force microscopy, AFM (Fig. 1c). These dispersions of isocyanate-derivatized graphite oxide allow graphene oxide sheets to be intimately mixed with many organic polymers, facilitating synthesis of graphene-polymer composites. Coupled with the possibility that graphite oxide can be chemically reduced (see above), we were hopeful that nanocomposites of polymer and isocyanate-derivatized graphite oxide could be rendered electrically conductive.

Hence, we were gratified to find that electrically conductive graphene-polymer nanocomposites can be prepared by solution-phase mixing of the exfoliated phenyl isocyanate-treated graphite oxide sheets with polystyrene, followed by their chemical reduction (Fig. 1d). These composites feature individual graphene sheets well dispersed throughout the polymer matrix (Fig. 1g). Similar dispersions have been achieved with other styrenic polymers such as acrylonitrile-butadiene-styrene and styrene-butadiene rubbers. Chemical reduction was essential for inducing electrical conductivity, as composite samples with un-reduced phenyl isocyanate-treated graphite oxide sheets were insulating. In addition, the presence of the polymer in solution during the reduction step was key to preventing the agglomeration of the sheets. As the reduction proceeds, the sheets become coated with the polymer and remain individually dispersed (Fig. 1d). When the reduction step preceded the introduction of the polymer, significant agglomeration of graphene

¹Department of Mechanical Engineering, ²Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3111, USA. ³School of Materials Engineering and Birk Nanotechnology Center, Purdue University, 501 Northwestern Avenue, West Lafayette, Indiana 47907, USA.

*These authors contributed equally to this work.

sheets occurred. Further details of the preparation and fabrication of composite samples are described in Methods.

In a concurrent research direction, we explored a rapid thermal expansion of graphite oxide ($\sim 1,000^\circ\text{C}$ under inert gas; R. K. Prud'homme *et al.*, manuscript in preparation). However, the thermally expanded graphite oxide platelets so produced are not fully exfoliated, consisting of several-nanometre-thick multilayer stacks, and their dispersion in a polymer matrix is a challenging step of manufacture. For the purpose of comparison, polystyrene composites prepared with such thermally expanded graphite oxide, by the same procedure as was used for phenyl isocyanate-treated graphite oxide, are less homogeneous and have higher electrical percolation thresholds (see Supplementary Information). For this and other reasons, we have focused on the bottom-up approach that yields molecular-level dispersion of individual graphene sheets as observed by scanning electron microscopy, SEM (Fig. 1g).

The quality of nanofiller dispersion in the polymer matrix directly correlates with its effectiveness for improving mechanical, electrical, thermal, impermeability and other properties. The properties of a composite are also intimately linked to the aspect ratio and surface-to-volume ratio of the filler. The potential properties of our graphene-

sheet-based composites thus appear promising owing to the extremely high aspect ratios of the sheets as determined from SEM images (Fig. 1g), in which the average lateral dimension was estimated to be $\sim 1\ \mu\text{m}$, similar to the values found by AFM (Fig. 1b, c). However, in contrast to the typically flat sheets seen by AFM when deposited onto atomically flat substrates, the sheets in the composites are crumpled and wrinkled, and at times folded.

At 2.4 volume per cent (vol.%) loading, the composite appears in the SEM images to be almost entirely filled with the graphene sheets, even though 97.6 vol.% is still filled by the polymer (Fig. 1g, right panel). This 'visual' effect is due to the enormous surface area of the sheets. The specific surface area of an individual graphene sheet is $\sim 2,600\ \text{m}^2\ \text{g}^{-1}$, so $1\ \mu\text{m}^3$ of a composite with 1 vol.% graphene sheets would have $\sim 50\ \mu\text{m}^2$ of sheet surface area. SEM image analysis was used to estimate the apparent graphene sheet surface area at different concentrations (Fig. 2a–d). It revealed a surface area lower by roughly a factor of two than the calculated values for each concentration. The nanoscale corrugation of the graphene sheets (not evident in SEM imaging) may contribute to the discrepancy. The intrinsic mechanics of the sheets, their crumpling, wrinkling and folding, and whether they can be processed to be non-crumpled (for example, so that

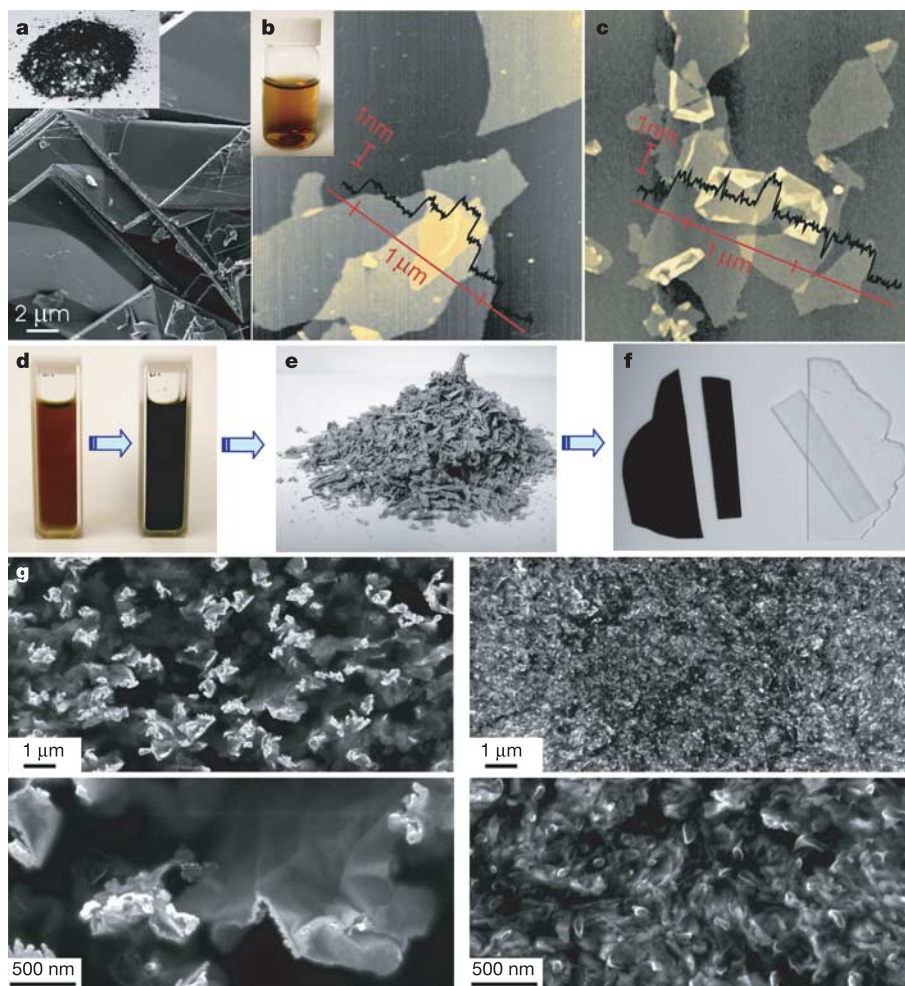


Figure 1 | Process flow of graphene-polymer composite fabrication. **a**, SEM and digital image (inset) of natural graphite. **b**, A typical AFM non-contact-mode image of graphite oxide sheets deposited onto a mica substrate from an aqueous dispersion (inset) with superimposed cross-section measurements taken along the red line indicating a sheet thickness of $\sim 1\ \text{nm}$. **c**, AFM image of phenyl isocyanate-treated graphite oxide sheets on mica and profile plot showing the $\sim 1\ \text{nm}$ thickness. **d**, Suspension of phenyl isocyanate-treated graphite oxide ($1\ \text{mg ml}^{-1}$) and dissolved polystyrene in

DMF before (left) and after (right) reduction by *N,N*-dimethylhydrazine. **e**, Composite powder as obtained after coagulation in methanol. **f**, Hot-pressed composite (0.12 vol.% of graphene) and pure polystyrene of the same 0.4-mm thickness and processed in the same way. **g**, Low (top row) and high (bottom row) magnification SEM images obtained from a fracture surface of composite samples of 0.48 vol.% (left) and 2.4 vol.% (right) graphene in polystyrene.

layering could be more effectively achieved) are issues for further study. Heat treatment of such composites to 300 °C, well above the glass transition temperature of polystyrene (100 °C), does not result in agglomeration of the graphene sheets (as assessed from extensive SEM imaging), suggesting their thermal stability over a broad temperature range.

As SEM cannot spatially resolve the thickness of an individual graphene-based sheet, transmission electron microscopy (TEM) was used to determine if the graphene-based sheets were indeed present in the composites as single, exfoliated sheets or as multi-layered platelets. Extensive high-resolution phase-contrast imaging showed no evidence of multi-layer stacks, even though SEM of the same slices clearly indicated the presence of nanoplatelets. In contrast, selected-area electron diffraction (SAED) of composite samples yielded spot patterns matching those expected for individual graphene-based sheets (Fig. 2e, f, and Supplementary Information). SAED of pure polystyrene samples fabricated in an identical manner (see Methods) showed the ring pattern expected for an amorphous material.

The conductive nature of the graphene filler motivated our study of the percolation behaviour through electrical measurements. As a control, we prepared polystyrene composites containing phenyl isocyanate-treated graphite oxide sheets that were not chemically reduced. At similar graphene concentrations, these samples were greyish in colour compared to the almost black composites filled with the reduced material, and they were not electrically conductive.

To obtain the true value of the conductivity of the composites, we made separate measurements of in-plane and transverse resistances

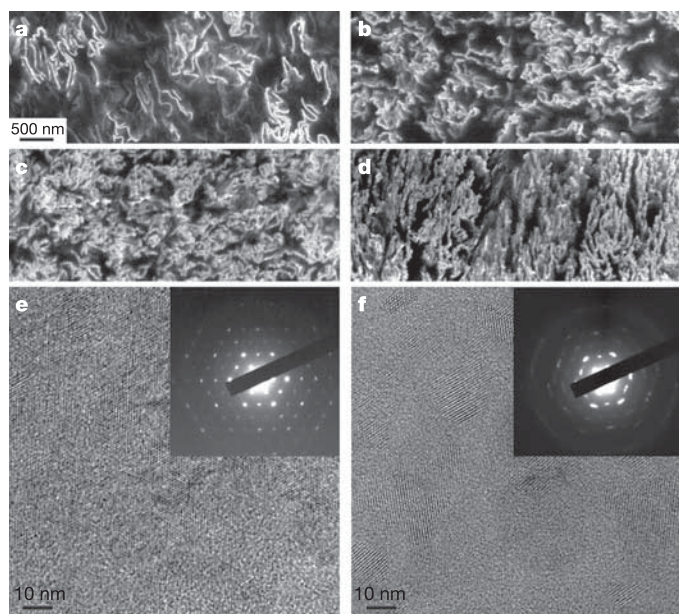


Figure 2 | SEM and TEM images of graphene-polystyrene composite. **a–d**, SEM images of the microtomed composites reveal different morphologies of the graphene sheets, including their packing, at different concentrations (vol.%): **a**, 0.24; **b**, 0.96; **c**, 1.44; and **d**, 2.4. Scale bar, shown in **a**, applies to **a–d**. **e**, **f**, High-resolution phase contrast images and SAED patterns (inset) of **e**, cast film made from powder composite, and **f**, microtomed composite sample. The SAED patterns show the six-fold rotational symmetry expected for diffraction with the beam incident along [0001]. Our experimentally obtained patterns show all of the reflections expected for such single graphene-based sheets. The *d* spacings associated with the spots were determined using both gold and graphite calibrants, and the first five sets of spots correspond to *d*-spacings (Å) of 4.23, 2.45, 2.12, 1.42 and 1.23. More details are presented in Supplementary Information. Additionally, these high-resolution images show regions where fringes are observed and regions where they are not, which indicates that there is significant local curvature in the sheets.

(Fig. 3, inset) and developed a numerical simulation program based on FEMLAB (version 3.1, Comsol AB), which takes into account geometry of the samples and electrical leads. All measurements and subsequent analysis revealed minor anisotropy, with the in-plane conductivity, $\sigma_{||}$, always slightly higher than the transverse conductivity, $\sigma_{\perp}$.

A rapid increase in the direct current electrical conductivity of composite materials takes place when the conductive filler forms an infinite network of connected paths through the insulating matrix. When the filler particles are rigid bodies, the conductivity of such media is typically described with a bond percolation model¹⁰. The conductivity of the composite, σ_c , above the percolation threshold is then treated with a power law: $\sigma_c = \sigma_f[(\phi - \phi_c)/(1 - \phi_c)]^t$ (ref. 24), where σ_f is the conductivity of the filler, ϕ the filler volume fraction, ϕ_c the percolation threshold (the onset of the transition), and *t* is the ‘universal critical exponent’.

Percolation in polystyrene–graphene composites occurs when the filler concentration, ϕ_c , is near 0.1 vol.% (Fig. 3). This percolation threshold is, to the best of our knowledge, about three times lower than reported for any other two-dimensional filler²⁵. Such a low percolation threshold for a three-dimensional isotropic case is evidently due to the extremely high aspect ratio of the graphene sheets and their excellent homogeneous dispersion in these composites. Randomly oriented oblate ellipsoids (disks) with an aspect ratio of 1,000 are predicted to have a geometric percolation threshold of ~ 0.1 vol.% (ref. 26). At a loading of about 0.15 vol.%, the conductivity of our composites already satisfies the antistatic criterion (10^{-6} S m^{-1}) for thin films¹² and rapidly rises over a 0.4 vol.% range. An increase in graphene sheet loading above 0.5 vol.% yields a more gradual increase in electrical conductivity, with values of $\sim 0.1 \text{ S m}^{-1}$ at 1 vol.% and $\sim 1 \text{ S m}^{-1}$ at 2.5 vol.%.

At present, research in nanocomposite materials using carbon-based nanofillers is dominated by carbon nanotubes. The electrical properties of our composites compare well with the best values reported in the literature for nanotube–polymer composites. The

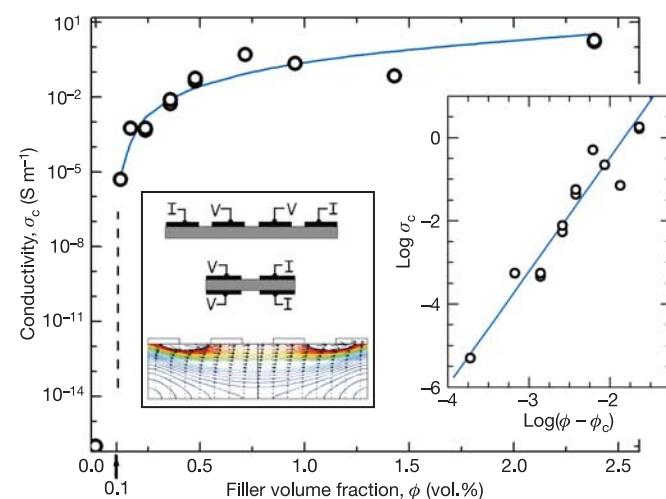


Figure 3 | Electrical conductivity of the polystyrene-graphene composites as a function of filler volume fraction. Main figure, composite conductivity, σ_c , plotted against filler volume fraction, ϕ . Right inset, $\log \sigma_c$ plotted against $\log(\phi - \phi_c)$, where ϕ_c is the percolation threshold (see text). Solid lines in both graphs are calculated conductivities based on the fitting (inset, log–log plot) of the experimental data to the effective conductivity equation described in the text. Fitted parameters are: $t = 2.74 \pm 0.20$, $\sigma_f = 10^{4.92 \pm 0.52} \text{ S m}^{-1}$ and $\phi_c = 0.1$ vol.%. Left inset: top and middle diagrams show the four-probe setup for in-plane and transverse measurements, respectively; bottom diagram, one of the computed distributions of the current density (contour lines) with local directions and magnitude (shown by arrows) in a specimen for the following conditions—the sample thickness is twice the electrode width and the gap between them, and the in-plane resistivity is 10 times lower than the transverse resistivity.

percolation threshold of around 0.1 vol.% is almost the same as was achieved for single-wall carbon nanotube (SWNT)–polystyrene composites made by latex technology²⁷, and just twice and four times higher than those measured for SWNT–polyimide^{11,24} and poly(phenyleneethynylene)-functionalized SWNT–polystyrene composites²⁸, respectively. Absolute conductivities of our graphene–polystyrene composites are also essentially the same as the values reported for SWNT-filled polystyrene composites²⁷. However, in contrast to these reports, which feature thin film composites, our samples are made by a quasi-industrial method (hot-pressing; we have also made samples by injection moulding that show similar values of electrical conductivity). Furthermore, graphene sheets have higher surface-to-volume ratios than SWNTs owing to the inaccessibility of the inner nanotube surface to polymer molecules. This makes graphene sheets potentially more favourable for altering all matrix properties—such as the mechanical, rheological and permeability properties, and degradation stability. Also, SWNTs are still much more expensive than graphite, which is a commodity material with about 1,000,000 t used annually, and which is sold for a few US dollars per kilogram²⁹. In contrast, graphite oxide can be readily prepared from a vast array of graphites, and its industrial production has also recently become a topic of interest³⁰.

We therefore expect that our method for preparation and incorporation of individual graphene sheets into polymer matrices will lead to the further development of a broad new class of materials with enhanced properties and even introduce new functionalities to polymer composites.

METHODS

Fabrication of composites. Graphite oxide was prepared by the Hummers method from SP-1 graphite (Bay Carbon), and dried for a week over phosphorus pentoxide in a vacuum desiccator. Dried graphite oxide (50 mg) was suspended in anhydrous DMF (5 ml, Dow-Grubbs solvent system), treated with phenyl isocyanate (2 mmol, Sigma-Aldrich) for 24 h, and recovered by filtration through a sintered glass funnel (50 ml, medium porosity). Stable dispersions of the resulting phenyl isocyanate-treated graphite oxide materials were prepared by ultrasonic exfoliation (Fisher Scientific FS60, 150 W, 1 h) in DMF (1 mg ml⁻¹). Polystyrene (Scientific Polymer Products, approximate $M_w = 280$ kD, PDI = 3.0) was added to these dispersions and dissolved with stirring (Fig. 1d, left). Reduction of the dispersed material (Fig. 1d, right) was carried out with dimethylhydrazine (0.1 ml in 10 ml of DMF, Sigma-Aldrich) at 80 °C for 24 h. Upon completion, the coagulation of the polymer composites was accomplished by adding the DMF solutions dropwise into a large volume of vigorously stirred methanol (10:1 with respect to the volume of DMF used). The coagulated composite powder (Fig. 1e) was isolated via filtration; washed with methanol (200 ml); dried at 130 °C under vacuum for ~10 h to remove residual solvent, anti-solvent, and moisture; crushed into a fine powder with a mortar and pestle, and then pressed (Fig. 1f) in a hydraulic hot press (Model 0230C-X1, PHI-Tulip) at 18 kN with a temperature of 210 °C. Further description of the composite samples' preparation is given in Supplementary Information. To convert wt% loading of graphene sheets in the composite samples to vol.% (as used in the text), a density for the phenyl isocyanate-treated graphite oxide sheets of 2.2 g cm⁻³ was assumed along with the known density of polystyrene, 1.05 g cm⁻³. Composites of acrylonitrile-butadiene-styrene and styrene-butadiene rubbers (Sigma-Aldrich) were similarly prepared using the appropriate copolymers.

Microscopy. AFM images were taken on an AutoProbe CP/MT scanning probe microscope (MultiTask; Veeco Instruments). Imaging was done in non-contact mode using a V-shaped 'Ultralever' probe B (Park Scientific Instruments, boron-doped Si with frequency $f_c = 78.6$ kHz, spring constants $k = 2.0$ – 3.8 N m⁻¹, and nominal tip radius 10 nm). All images were collected under ambient conditions at 50% relative humidity and 23 °C with a scanning raster rate of 1 Hz. Samples for AFM images were prepared by depositing dispersions of functionalized graphite oxide in DMF on a freshly cleaved mica surface (Ted Pella Inc.) and allowing them to dry in air.

Electron micrographs were acquired using a field emission SEM (LEO 1525) and a TEM (JEM-3010, JEOL, operated at 300 kV). Two SEM imaging conditions were used: 'surface imaging' operating at ≤ 1 kV, and 'sub-surface imaging' at ≥ 5 kV. Microtomed and cast film samples were prepared for TEM. Pressed composites were microtomed to slices of 40–60 nm thickness (Leica Ultracut UCT, Reichert Inc.) and dropped onto standard TEM grids.

Samples investigated in this way were 0.24, 1.44 and 2.4 vol.%. A polystyrene control sample was also prepared and imaged in this way. Before imaging with TEM, SEM was used to verify that a sufficient amount of highly dispersed platelets were present. Cast-film samples were prepared to image graphene sheets 'lying flat' on the carbon support film, with the beam incident along the [0001] direction of the graphene sheet. A small amount (<0.01 mg) of un-pressed powder composite was dissolved in toluene (10 ml) together with additional polystyrene (<0.02 mg) as diluent. The solution was sonicated (Crest Ultrasonic Tru-Sweep 175HT, 45 W, 1 h) before casting, producing a highly dilute solution that, when cast on a mica substrate, formed a film approximately 40 nm thick, as measured by AFM.

Electrical measurements. Hot-pressed composite samples having 0.3–0.5 mm thickness were cut into strips (1–2 mm wide and 15–20 mm long). A thin 'skin' layer deficient in the phenyl isocyanate-treated graphite oxide platelets was removed by oxygen plasma etching (Plasma-Preen II-862, Plasmatic Systems; 2 torr O₂, 1.5 min, 350 W) to reduce the contact resistance between the metal leads and sample tested. After etching, a thin gold film (~25 nm) was thermally deposited (BOC Edwards Auto 306 Evaporation System) on one side of the sample for 'longitudinal' measurements (gap between contacts is 0.15 mm), and on the top and bottom sides for 'transverse' measurements. The d.c. resistance of the composites was measured with a d.c. power supply (HP6612C), digital multimeter (HP34401A) and picoammeter (Keithley 6485) using a standard four-probe technique with a threshold detection limit of 0.1 GΩ.

Received 18 January; accepted 5 June 2006.

- Dresselhaus, M. S. & Dresselhaus, G. Intercalation compounds of graphite. *Adv. Phys.* **51**, 1–186 (2002).
- Hirata, M., Gotou, T., Horiuchi, S., Fujiwara, M. & Ohba, M. Thin-film particles of graphite oxide 1: High-yield synthesis and flexibility of the particles. *Carbon* **42**, 2929–2937 (2004).
- Yu, M. F., Lourie, O., Moloni, K., Kelly, T. F. & Ruoff, R. S. Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load. *Science* **287**, 637–640 (2000).
- Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
- Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
- Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
- Zhang, Y., Small, J. P., Amori, M. E. S. & Kim, P. Electric field modulation of galvanomagnetic properties of mesoscopic graphite. *Phys. Rev. Lett.* **94**, 176803 (2005).
- Berger, C. *et al.* Ultrathin epitaxial graphite: two-dimensional electron gas properties and a route toward graphene-based nanoelectronics. *J. Phys. Chem. B* **108**, 19912–19916 (2004).
- Stankovich, S. *et al.* Stable aqueous dispersions of graphitic nanoplatelets via the reduction of exfoliated graphite oxide in the presence of poly(sodium 4-styrenesulfonate). *J. Mater. Chem.* **16**, 155–158 (2006).
- Stauffer, D. & Aharony, A. *Introduction to Percolation Theory* (Taylor & Francis, London, 1991).
- Ounaies, Z., Park, C., Wise, K. E., Siochi, E. J. & Harrison, J. S. Electrical properties of single wall carbon nanotube reinforced polyimide composites. *Compos. Sci. Technol.* **63**, 1637–1646 (2003).
- Chung, D. D. L. Electrical applications of carbon materials. *J. Mater. Sci.* **39**, 2645–2661 (2004).
- He, H., Klinowski, J., Forster, M. & Lerf, A. A new structural model for graphite oxide. *Chem. Phys. Lett.* **287**, 53–56 (1998).
- Lerf, A., He, H., Forster, M. & Klinowski, J. Structure of graphite oxide revisited. *J. Phys. Chem. B* **102**, 4477–4482 (1998).
- Hirata, M., Gotou, T. & Ohba, M. Thin-film particles of graphite oxide 2: Preliminary studies for internal micro fabrication of single particle and carbonaceous electronic circuits. *Carbon* **43**, 503–510 (2005).
- Szabo, T., Szeri, A. & Dekany, I. Composite graphitic nanolayers prepared by self-assembly between finely dispersed graphite oxide and a cationic polymer. *Carbon* **43**, 87–94 (2005).
- Kovtyukhova, N. I. *et al.* Layer-by-layer assembly of ultrathin composite films from micron-sized graphite oxide sheets and polycations. *Chem. Mater.* **11**, 771–778 (1999).
- Kotov, N. A., Dekany, I. & Fendler, J. H. Ultrathin graphite oxide-polyelectrolyte composites prepared by self-assembly. Transition between conductive and non-conductive states. *Adv. Mater.* **8**, 637–641 (1996).
- Cassagneau, T., Guerin, F. & Fendler, J. H. Preparation and characterization of ultrathin films layer-by-layer self-assembled from graphite oxide nanoplatelets and polymers. *Langmuir* **16**, 7318–7324 (2000).
- Vaia, R. A. & Wagner, H. D. Framework for nanocomposites. *Mater. Today* **7**, 32–37 (2004).
- Du, X. S., Xiao, M., Meng, Y. Z. & Hay, A. S. Novel synthesis of conductive poly(arylene disulfide)/graphite nanocomposite. *Synth. Met.* **143**, 129–132 (2004).

22. Liu, P. & Gong, K. Synthesis of polyaniline-intercalated graphite oxide by an in situ oxidative polymerization reaction. *Carbon* **37**, 706–707 (1999).
23. Stankovich, S., Piner, R. D., Nguyen, S. T. & Ruoff, R. S. Synthesis and exfoliation of isocyanate-treated graphene oxide nanoplatelets. *Carbon* (in the press).
24. McLachlan, D. S. *et al.* AC and DC percolative conductivity of single wall carbon nanotube polymer composites. *J. Polym. Sci. B* **43**, 3273–3287 (2005).
25. Chen, G., Weng, W., Wu, D. & Wu, C. PMMA/graphite nanosheets composite and its conducting properties. *Eur. Polym. J.* **39**, 2329–2335 (2003).
26. Garboczi, E. J., Snyder, K. A., Douglas, J. F. & Thorpe, M. F. Geometrical percolation threshold of overlapping ellipsoids. *Phys. Rev. E* **52**, 819–828 (1995).
27. Grossiord, N., Loos, J. & Koning, C. E. Strategies for dispersing carbon nanotubes in highly viscous polymers. *J. Mater. Chem.* **15**, 2349–2352 (2005).
28. Ramasubramaniam, R., Chen, J. & Liu, H. Homogeneous carbon nanotube polymer composites for electrical applications. *Appl. Phys. Lett.* **83**, 2928–2930 (2003).
29. Olson, D. W. USGS Mineral Commodity Summaries. (<http://minerals.usgs.gov/minerals/pubs/commodity/graphite/graphmcs06.pdf>) (2006).
30. Titelman, G. I. & Bron, S. *Method of Manufacturing of Graphite Oxide (GO)*

P1391–P1392 (Research Disclosure 500, IMI TAMI Institute for Research & Development Ltd., Haifa, 2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements SEM was done at the Electron Probe Instrumentation Centre at Northwestern University. This work was funded by the NASA University Research, Engineering and Technology Institute on Bio-Inspired Materials (BIMat) and by the NSF NIRT grant 'Nanostructured carbons from self-assembled block copolymer precursors: From synthesis and characterization to devices'. We appreciate J. A. Ibers and A. L. Ruoff for critically reading an early version of this manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S.R. (r-ruoff@northwestern.edu) or S.T.N. (stn@northwestern.edu).

Seismic reflection images of the Moho underlying melt sills at the East Pacific Rise

S. C. Singh^{1,2}, A. J. Harding³, G. M. Kent³, M. C. Sinha⁴, V. Combier¹, S. Bazin¹, C. H. Tong², J. W. Pye², P. J. Barton², R. W. Hobbs⁵, R. S. White² & J. A. Orcutt³

The determination of melt distribution in the crust and the nature of the crust–mantle boundary (the ‘Moho’) is fundamental to the understanding of crustal accretion processes at oceanic spreading centres. Upper-crustal magma chambers have been imaged beneath fast- and intermediate-spreading centres^{1–4} but it has been difficult to image structures beneath these magma sills. Using three-dimensional seismic reflection images, here we report the presence of Moho reflections beneath a crustal magma chamber at the 9° 03′ N overlapping spreading centre, East Pacific Rise. Our observations highlight the formation of the Moho at zero-aged crust. Over a distance of less than 7 km along the ridge crest, a rapid increase in two-way travel time of seismic waves between the magma chamber and Moho reflections is observed, which we suggest is due to a melt anomaly in the lower crust. The amplitude versus offset variation of reflections from the magma chamber shows a coincident region of higher melt fraction overlying this anomalous region, supporting the conclusion of additional melt at depth.

Multichannel seismic (MCS) reflection studies from fast- and intermediate-spreading centres have shown the presence of upper-crustal melt sills, which has led to the suggestion that melt accumulates predominantly within and immediately below these features^{1–4}. These melt sills are typically 250–4,000 m wide, <100 m thick^{5,6} and are underlain by a broad, partially molten region characterized by low seismic velocities^{7–10} and high attenuation¹¹. Some seismic and seafloor compliance studies suggest the presence of melt within or beneath the Moho—the transition zone separating lower crust and upper mantle—in the form of a melt sill or a large partially molten region^{12–14}. The detailed nature of upper-crustal magma chambers is now well understood^{4,15}, but the geometry and size of lower-crustal melt bodies and their relationship to subcrustal reservoirs is still poorly known.

In 1997, a three-dimensional (3D) MCS reflection and tomography experiment was carried out over the 9° 03′ N overlapping spreading centre (OSC) at the East Pacific Rise (EPR). The 3D MCS data were acquired over a 20 km by 20 km area of sea floor (Fig. 1) covering both limbs of the OSC. The 9° 03′ N OSC has an offset of 8 km and the two limbs overlap for about 27 km. The eastern limb of the OSC is propagating southward at a rate of 42 km Myr^{−1}, or about half of the present-day spreading rate, whereas the western limb is receding¹⁶.

The extensive presence of axial magma chamber (AMC) reflections beneath both limbs and portions of the overlap basin⁴ is indicative of a robust supply of magma beneath this feature (Fig. 1 inset). The AMC reflections located at upper-crustal depths between ~1,650 and 2,200 m beneath the sea floor are up to 4.5 km wide in the north, narrowing down to ~250 m in the south. On the eastern limb,

the melt sill is asymmetric and lies mainly west of the ridge crest. The geometry of these magma sills is not consistent with large-scale crustal migration of melt away from mantle upwelling at the segment centre, but instead seems to be related to vertical transport from the underlying mantle and subsequent focusing in the mid-crust.

Using 3D MCS reflection data processed in a new manner to

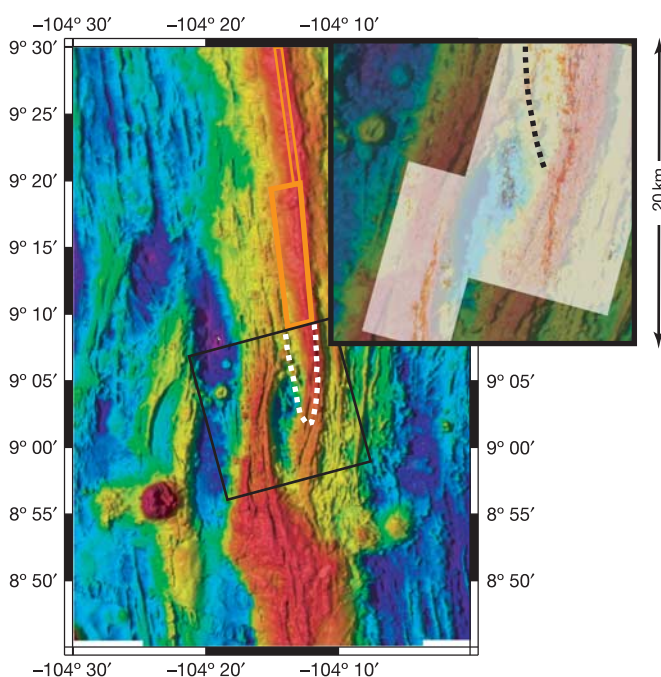


Figure 1 | Three-dimensional seismic survey area and images of melt sills. Seafloor bathymetry along the 9° N segment of the East Pacific Rise spreading centre. The black square at 9° N shows the location of the 3D seismic survey (Anatomy of a Ridge Axis Discontinuity, ARAD), positioned across the OSC. The 3D survey size is 20 km × 20 km, and is oriented 15° anticlockwise from the north. Location of the anomalous magma-chamber-to-the-Moho travel time is shown (white dashed line) and possible extension northwards up to 9° 17′ N (orange lines). The inset shows a top-down view of magma-chamber structure seen beneath a semi-transparent sea floor centred at the 9° 03′ N overlap basin. The underlying 3D seismic reflectivity volume has been rendered using an opacity function to highlight relationships between detailed seafloor structure and magma-chamber geometry. Subtle hues of orange and blue amplitudes reveal the location of strong subsurface AMC reflections. The dashed black line delineates the westward extension of the magma sill beneath the northern overlap basin, which is coincident with a small volcanic ridge.

¹Laboratoire de Géosciences Marines, IPG Paris, 75252 Paris cedex 05, France. ²Bullard Laboratories, University of Cambridge, Cambridge CB3 0EZ, UK. ³Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, California 92093, USA. ⁴School of Ocean and Earth Sciences, University of Southampton, Southampton SO14 3ZH, UK. ⁵Department of Earth Sciences, University of Durham, Durham DH1 3LE, UK.

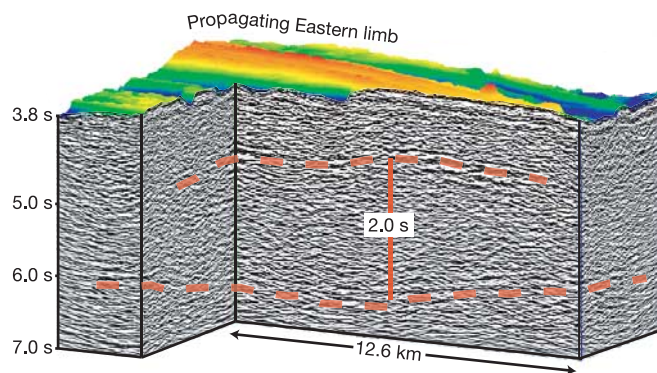


Figure 2 | Three-dimensional view of reflection images of the Moho and melt sill. A 3D cut through the ARAD seismic reflectivity volume revealing the observed travel-time anomaly (~ 2.1 s maximum value) between the AMC and the base of the crust (the Moho). Multibeam bathymetry has been included to provide spatial context between seafloor morphology and underlying crustal structure. This 3D cube spans an area extending $15.5 \text{ km} \times 11.5 \text{ km}$ covering most of the eastern limb and overlap basin. Part of the volume has been removed to show the relationship between the sea floor, the AMC and the Moho (highlighted) reflections. See Fig. 3a for the location of the trimmed 3D volume. AMC reflections are imaged along both overlap limbs and beneath the northern half of the overlap basin, whereas Moho reflections are present throughout most of the 3D volume (although absent beneath the western limb). Note a rapid increase in two-way travel time between the AMC and Moho reflections along the displayed 12.6-km-long facet of the volume (increasing from 1.3 s to 2.0 s over a distance spanning only $<7 \text{ km}$).

highlight deeper structure, we observed Moho reflections beneath the AMC along the entire eastern limb (Figs 2 and 3a). The polarity of the Moho reflections is complex, but it is generally positive, similar to the seabed reflections, requiring a positive velocity contrast. The existence of near-vertical incidence reflections from the Moho means that the crust–mantle boundary is sharp compared to the wavelength (300 m) of the seismic energy at this depth. This critical relationship has never been observed on any two-dimensional (2D) MCS profile; thus, these new 3D data now permit the exploration of changes in lower crustal structure between the magma chamber and Moho boundary.

A part of the 3D volume of reflection data has been removed to focus on crustal structure beneath the eastern limb in Fig. 2 and Supplementary Fig. S2, highlighting the relationship between AMC and Moho reflections. Moho reflections are easily identifiable on the along-axis sections of the 3D seismic volume, although they are also present on the cross-axis sections (Fig. 2 and Supplementary Figs S1 and S2). Moho reflections are observed over 50% of the 3D box; these events are predominantly seen beneath the eastern limb, the northern part of the overlap basin and west of the receding limb (Fig. 3a). There could be significant attenuation¹¹ within the upper-crustal magma chamber ($Q_p \approx 10\text{--}20$), so a strong impedance contrast between the uppermost mantle and lower crust would be needed to produce the observed Moho reflections. The compressional-wave (P-wave) velocity of the solid mantle beneath the ridge axis is about $7.6\text{--}7.8 \text{ km s}^{-1}$. It is therefore likely that a low P-wave velocity (for example, $5.0\text{--}6.0 \text{ km s}^{-1}$) in the lower crust owing to additional melt would be needed to produce a visible reflection; if there exists reduced mantle velocity due to subcrustal melt accumulation, this effect would be offset, requiring even lower velocity above the Moho.

There are two possibilities for the presence of the Moho beneath the AMC: (1) the AMC has propagated over the pre-existing Moho underlying older crust, (2) the Moho was formed at zero age. Because Moho reflections are present beneath a wide AMC (4.5 km) in the north along the propagating limb of the spreading ridge, we prefer the latter explanation for the presence of the Moho. This observation

suggests that differentiation of oceanic crust occurs over a very limited spatial region, and hence, a short period of geologic time; although this fundamental boundary may be modified as crust slides away from the rise axis, these new images reveal insights into the dynamics of the lower-crustal formation.

Off-axis, the Moho reflections are observed between 1.9 and 2.3 s below the sea floor (Fig. 3a), which are similar to those observed in ref. 17 using widely spaced 2D seismic lines, whereas these reflections are 2.6–2.9 s below the sea floor beneath the wide AMC on the eastern limb, about 0.6–0.7 s more than the off-axis travel time. If we assume an average crustal velocity of 6.0 km s^{-1} , the crust would be about 6–7 km thick off-axis, which is consistent with the crustal thickness determined from tomographic study outside the 3D box¹⁸. However, the crust would be around 8–9 km thick beneath the AMC on the eastern limb, which is anomalous.

It is intriguing to note that the contribution to this large travel-time anomaly comes from the difference in travel time between the AMC and Moho reflections, that is, from the lower crust. For example, the two-way travel time between the AMC and Moho reflections is about 1.30 s near the rift tip in the south, but then increases dramatically to over 2.0 s along the flank of the eastern ridge abutting the northern overlap basin, within a distance of $<7 \text{ km}$ along the axis and $<5 \text{ km}$ across the axis; that is, a short wavelength increase in travel-time difference of 35–40% in the lower crust. If the whole anomaly is produced by the presence of melt then over 45% of the lower crust could be molten, assuming horizontally aligned, spheroidal inclusions with an aspect ratio of 1/10 between the semi-minor and semi-major axes¹⁹.

This lower-crustal travel-time anomaly could be caused either by a thick lower crust, the presence of a large amount of melt, or a combination of both. If we assume that the average P-wave velocity in the lower crust is 6.8 km s^{-1} at the ridge axis, where the temperature should be around $1,100^\circ\text{C}$, then the thickness of the lower crust below the propagating rift tip would be 4.4 km, in contrast to just 7 km further north, where the lower-crustal thickness value would exceed 6.8 km. Under this scenario, an unexpected crustal thickness variation from 6.2 to 8.5 km would occur over a very short distance (5–7 km), which has never been observed before. Canales *et al.*¹⁸ found a V-shaped wake located about 20 km north of the OSC where the crust is thick (7.3–7.8 km). Therefore, crustal thickening can explain only a part of the total signal observed, and a significant amount of melt in the lower crust would be required to explain the large travel-time anomaly beneath the AMC. Furthermore, to create a thicker crust, extra melt supply would be required, supporting the idea that there must be a large amount of melt in the lower crust.

To gain more insight into the source of this lower-crustal travel-time anomaly, the reflection volume was reprocessed using a novel technique¹⁵ to investigate the internal structure of the overlying AMC. Amplitude variation with offset (AVO) studies can be used¹⁵ to distinguish melt-rich regions with fresh magma supply from areas of only partial melt that have been cooled and crystallized. The 3D AVO results (Fig. 3d) reveal that the northern OSC basin and eastern limb are underlain by a melt-rich sill, while the southern most reaches of the propagator tip are melt-starved or crystal-rich. This geographic association of melt-rich region with lower-crustal anomaly suggests that the melt-rich sill may be replenished by a melt anomaly within the lower crust.

Tomographic studies²⁰ also show a broad low-velocity anomaly at mid-crustal levels (3 km below the sea floor), which is roughly coincident with the overlying melt-rich sill and lower-crustal anomaly. The shallowest mantle is also characterized by lower velocities beneath this site²¹. Similarly, seafloor compliance measurements show very low S-wave velocities in the lower crust near this anomalous site²². Taken together, these various measurements provide evidence for a melt-rich delivery system at all depth levels, where melt is transported across the mantle into the lower crust, stored within the lower crust and the upper-crustal magma sill.

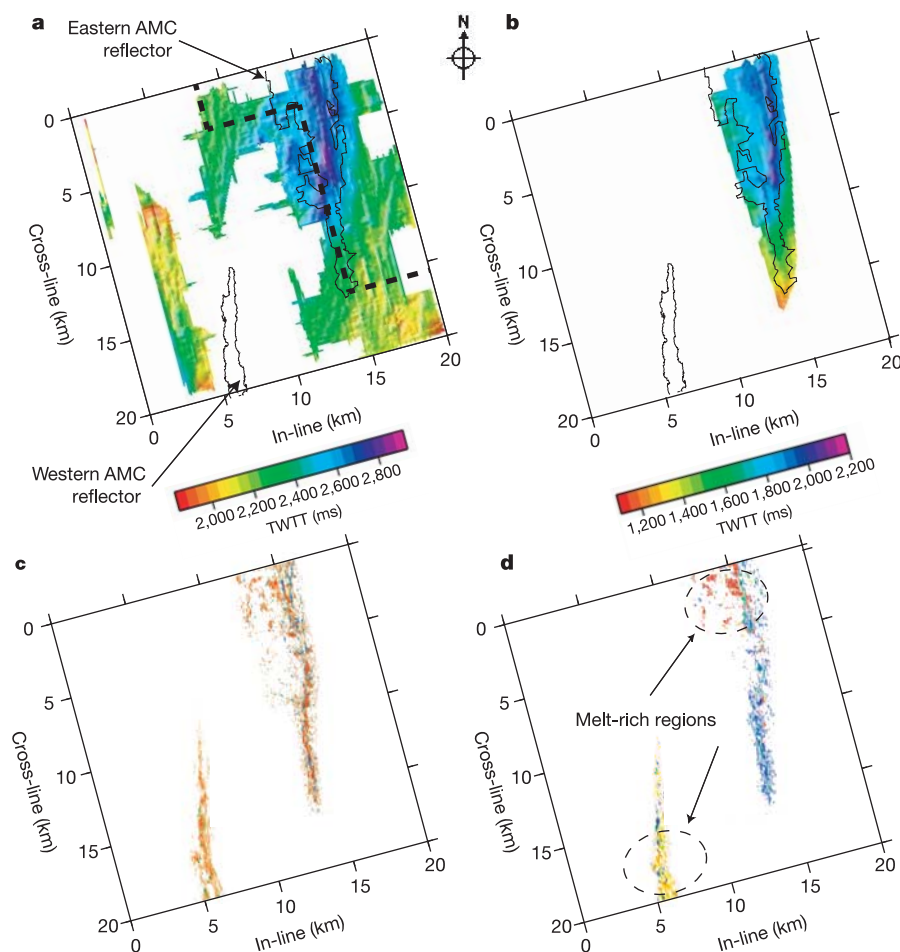


Figure 3 | Two-way reflection times of Moho and amplitude versus offset for melt sill. **a, b,** Two-way reflection travel times (TWTT) of the Moho to the sea floor (**a**) and the Moho to the AMC (**b**). Spatial outline of magma chamber reflections are shown as thin black lines (**a** and **b**). The geographical position of the 3D subvolume shown in Fig. 2 is marked by a thick dashed line (**a**). The maximum Moho to AMC travel time is nearly 2.1 s (**b**), located near the transition where the AMC narrows and plunges down the propagating eastern ridge tip⁴. **c, d,** AMC reflection amplitude map using all streamer source–receiver offsets (red–orange–green–blue colour palette; **c**) and a composite of the inner half (red–orange) and the outer half (blue–green) of streamer (**d**). **d,** The AVO analysis highlights regions within the survey area that are melt-rich. These melt-laden regions lack any significant transmission of S-waves through the magma sill, resulting in a characteristic drop-off of P-wave energy in the 1.6–3.2-km offset band¹⁵. The AVO amplitude map shows a reflection image (blue–green) using the outer half of the streamer, combined on top of the image from the inner half (red–orange); in regions where S-wave transmission through the sill occurs (where the sill is crystal-rich), significant reflected energy in the outer range band (blue–green) masks the red–orange image from more vertically travelling energy. Melt-rich regions (red–orange) are found just to the north and south of the overlap basin centre; the melt-rich region in the northern half of the survey area corresponds with the underlying Moho–AMC travel time anomaly, suggesting that increased melt within the lower crust may be important in the observed travel-time delay.

The southern limit of the lower-crustal anomaly is marked by a thin lower crust (1.3 s), with a total crustal thickness of ~ 6 km near the southern tip of the eastern limb (Fig. 3b). Beneath the eastern limb, it follows the asymmetric AMC reflections and seafloor morphology, and is elongated in the N15°W direction (Figs 1 and 3b). Kent *et al.*⁵ have observed a wide (4 km) asymmetric melt sill south of the 9° 17' N ridge-axis discontinuity. The depth and nature of this extended melt sill is similar to the AMC observed along the eastern limb of the OSC. If the presence of the Moho and melt anomaly is associated with the wide melt sill, then this lower-crustal anomaly may continue up to 9° 17' N. The presence of young basalt²³ (< 50 yr) between 9° 12' N and 9° 14' N and black smokers²⁴ near 9° 16.8' N might be associated with the large fresh melt supply in the crust from the mantle⁶.

The 3D data volume can also be used to search for large melt sills within the lower crust. The seismic imaging condition and the continuity of the Moho reflections allow us to test whether similar reflections from melt-sill events exist between the AMC and the crust–mantle boundary to verify whether the lower crust might be formed by a series of sill intrusions within the lower crust^{25,26}. The dominant frequency used in our study is 10–30 Hz, and the dominant wavelength at these depths is about 300–600 m, assuming a P-wave velocity of 6.8 km s^{-1} . Therefore, the resolution in our data is about 500 m laterally, and 50 m in depth; any melt sill with dimensions greater than this volume should be imaged. But no laterally continuous events are observed between the Moho and AMC reflections in this data volume, so we can confidently say that there are no high-amplitude, large seismic events between the AMC and the Moho that are of similar size when compared to the melt sills on top of the low-velocity zone. Melt sills of smaller size could be present, but we cannot resolve them using our present data set. The

absence of any observable melt sills in the lower crust suggests that the melt in the lower crust could be distributed in small pores or in very small pockets.

These results are consistent with a model where the lower crust is porous on an intergranular scale, and melt is efficiently transported to the shallower melt sill²⁷. Our results are also consistent with crustal accretion models where crystallization takes place in very small melt pockets at a range of depths within the axial crust²⁸. Using the correlation between the seafloor morphology along the ridge axis and ridge migration relative to a fixed hotspot frame of reference, Carbotte *et al.*²⁹ suggest that the variations in the seafloor morphology may result from melt focusing across discontinuities from an asymmetric zone of mantle upwelling beneath the advancing plate. The large melt anomaly in the crust at the northern end of the OSC might have resulted from such an asymmetric mantle upwelling. These results lead us to suggest that the localized asymmetric large lower-crustal melt anomaly could be responsible for the propagation of the OSC southwards.

METHODS

Navigation, binning and imaging. The 3D seismic reflection data were acquired within a 20 km by 20 km patch of sea floor (Fig. 1) positioned over the OSC at 9° 03' N, using the academic research vessel RV *Maurice Ewing*. In this box, over 201 seismic profiles were collected at 100-m intervals using a 3,100-m-long streamer, with an in-line shot spacing of 38 m. The shooting direction was rotated 15° counter-clockwise of north, which was roughly perpendicular to the ridge axis.

In-line 2D profiles were corrected for dip-moveout, using water velocity ($1,500 \text{ m s}^{-1}$) before 3D binning. Subsequently, these data were re-gathered in 12.5 m by 100 m common-midpoint (CMP) 3D bins, with an elastic binning strategy extending out 50 m on either side of the active bin (to even out offset coverage within individual CMP gathers)³⁰. Velocity analysis of CMP gathers was performed every 100 m along every 10th in-line profile (or 1.0 km cross-line

spacing). To enhance deeper events such as Moho reflections, only source–receiver offsets up to 2,300 m were stacked down to 5.0 s two-way travel time, with greater ranges included up to 3,100 m below this time horizon in a tapered fashion. This muting strategy was used to minimize strong reverberations within layer 2A, which can potentially mask weaker, underlying Moho reflections. Seafloor multiples were digitized and muted to avoid artefacts during the migration or imaging sequence.

A two-pass Kirchhoff time migration was performed using 95% of the stacking velocity. First, a two-dimensional in-line migration operation was performed, which allowed the focusing of energy in the up-dip in-line direction. To reduce the spatial aliasing effect in the cross-line direction, the data were then interpolated onto a 25 m by 25 m grid. Finally, a two-dimensional cross-line Kirchhoff time migration was applied to focus energy up-dip in the cross-line direction. This processing sequence is significantly different in several respects from that of ref. 4, which was focused on enhancing signals for layer 2A and magma-chamber reflections, whereas 2D dip-moveout, far-offset muting and Kirchhoff migration, when combined, provided a better low-frequency imaging tool for deeper reflections such as the Moho.

Amplitude variation with offset analysis. The AVO analysis was performed using the same basic processing scheme presented by ref. 4, but modified to select specific source–receiver ranges to produce complementary range-gated images. Following the methodology of ref. 15, energy reflecting from the magma sill was processed in a parallel fashion for source–receiver offsets below and above ~1,625 m, which effectively isolates reflections from a mushy sill on the outer half of the streamer. For a melt-rich sill, the inability to transmit shear-wave (S-wave) energy across this body results in a precipitous drop in P-wave reflection amplitude, with minimal values in a range band extending from ~2 to 3 km. Conversely, a crystal-rich sill capable of supporting S-wave propagation is characterized by only a slight decrease in reflection amplitudes for ranges extending from ~0 to 3 km.

Next, these two AVO images can be combined to highlight areas of the magma sill that can or cannot transmit S-wave energy (or melt-rich or mushy); for example, if the image from the outer half of the streamer is layered on top of the inner half, wherever the magma body is crystal-rich, that particular image (outer half) will essentially mask energy from the inner half of the streamer. Correspondingly, for a melt-rich sill, there will be very little energy in this outer band, so only the inner-half image is visible. With an appropriate use of complementary colour palettes for both images, variations in magma sill properties can easily be shown.

Received 15 February; accepted 24 May 2006.

- Detrick, R. S. *et al.* Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
- Mutter, J. C. *et al.* Magma distribution across ridge axis discontinuities on the East Pacific Rise (8°40' to 9°50' North) from multi-channel seismic images. *Nature* **336**, 156–158 (1988).
- Collier, J. & Sinha, M. C. Seismic images of a magma chamber beneath the Lau basin back-arc spreading centre. *Nature* **346**, 646–648 (1990).
- Kent, G. M. *et al.* Evidence from three-dimensional reflectivity images for enhanced melt supply beneath mid-ocean-ridge discontinuities. *Nature* **406**, 614–618 (2000).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. Distribution of magma beneath the East Pacific Rise between the Clipperton Transform and the 9° 17' N Deval from forward modelling of common depth point data. *J. Geophys. Res.* **98**, 13945–13969 (1993).
- Singh, S. C., Collier, J. S., Kent, G. M., Harding, A. J. & Orcutt, J. A. Seismic evidence for a hydrothermal layer above the solid roof of axial magma chamber at the southern East Pacific Rise. *Geology* **27**, 219–222 (1999).
- Harding, A. J. *et al.* The structure of young oceanic crust at 13° N on the East Pacific Rise from expanding spread profiles. *J. Geophys. Res.* **94**, 12163–12196 (1989).
- Vera, E. E. *et al.* The structure of 0 to 0.2 m.y. old oceanic crust at 9° N on the East Pacific Rise from expanding spread profiles. *J. Geophys. Res.* **95**, 15529–15556 (1990).
- Toomey, D. R., Purdy, G. M., Solomon, S. C. & Wilcock, W. S. D. The three-dimensional seismic velocity structure of the East Pacific Rise near latitude 9°30' N. *Nature* **347**, 639–645 (1990).
- Dunn, R. A., Toomey, D. R. & Solomon, S. C. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9°30' N. *J. Geophys. Res.* **105**, 23537–23555 (2000).
- Wilcock, W. S. *et al.* Seismic attenuation structure of the East Pacific Rise near 9° 30' N. *J. Geophys. Res.* **100**, 24147–24165 (1995).
- Dunn, R. A. & Toomey, D. R. Seismological evidence for three-dimensional melt migration beneath the East Pacific Rise. *Nature* **388**, 259–262 (1997).
- Crawford, W. C., Webb, S. C. & Hilderbrand, J. A. Constraints on melt in the lower crust and Moho at the East Pacific Rise, 9°48' N, using seafloor compliance measurements. *J. Geophys. Res.* **104**, 2923–2939 (1999).
- Garmann, J. Accumulations of melt at the base of young oceanic crust. *Nature* **340**, 628–632 (1989).
- Singh, S. C., Kent, G. M., Collier, J. S., Harding, A. J. & Orcutt, J. A. Melt to mush variations in crustal magma properties along the ridge crest at the southern East Pacific Rise. *Nature* **394**, 874–878 (1998).
- Carbotte, S. & Macdonald, K. C. Evolution of ridge segments and discontinuities from SeaMARC II and three-dimensional magnetic studies. *J. Geophys. Res.* **97**, 6959–6982 (1992).
- Barth, G. A. & Mutter, J. C. Variability in oceanic crustal thickness and structure: Multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101**, 17951–17975 (1996).
- Canales, J.-P. *et al.* Segment-scale variations in the crustal structure of 150–300 kyr old fast spreading oceanic crust (East Pacific Rise, 8 15 N–10 5 N) from wide-angle seismic refraction profiles. *Geophys. J. Int.* **152**, 766–794 (2003).
- Taylor, M. A. & Singh, S. C. Composition and microstructure of magma bodies from effective medium theory. *Geophys. J. Int.* **149**, 15–21 (2002).
- Bazin, S. *et al.* A three-dimensional study of axial low velocity region beneath the 9°03' overlapping spreading center. *Geophys. Res. Lett.* **30**, 11–14, doi:10.1029/2002GL015137 (2003).
- Dunn, R. A., Toomey, D. R., Detrick, R. S. & Wilcock, W. S. D. Continuous mantle melt supply beneath an overlapping spreading center on the East Pacific Rise. *Science* **291**, 1955–1958 (2001).
- Crawford, W. C. & Webb, S. C. Variations in the distribution of magma in the lower crust and the Moho beneath a fast spreading mid-ocean ridge. *Earth Planet. Sci. Lett.* **203**, 117–130 (2002).
- Wright, D., Haymon, R. M. & Fornari, D. J. Crustal fissuring and its relationship to magmatic and hydrothermal processes on the East Pacific Rise crest (9°12' to 54° N). *J. Geophys. Res.* **100**, 6097–6120 (1995).
- Haymon, R. M. *et al.* Hydrothermal vent distribution along the East Pacific Rise crest (9°9'–54° N) and its relationship to magmatic and tectonic processes on fast-spreading mid-ocean ridges. *Earth Planet. Sci. Lett.* **104**, 513–534 (1991).
- Boudier, F., Nicolas, A. & Idefonse, B. Magma chambers in the Oman Ophiolite: fed from the top and bottom. *Earth Planet. Sci. Lett.* **144**, 239–250 (1996).
- Kelemen, P. B., Koga, K. & Shimizu, N. Geochemistry of gabbro sills in the crust-mantle transition zone of the Oman Ophiolite; implications for the origin of the oceanic lower crust. *Earth Planet. Sci. Lett.* **146**, 475–488 (1997).
- Nicolas, A. *Structure of Ophiolites and Dynamics of Oceanic Lithosphere* (Kluwer, Amsterdam, 1989).
- MacLennan, J. A., Hulme, T. & Singh, S. C. Thermal models of oceanic crustal accretion: Linking geophysical, geological and petrological observations. *Geophys. Geochem. Geosyst.* **5**, doi:10.1029/2003GC000605 (2003).
- Carbotte, S. M., Small, C. & Donnelly, K. The influence of ridge migration on the magmatic segmentation of mid-ocean ridges. *Nature* **429**, 743–746 (2004).
- Yilmaz, O. *Seismic Data Analysis* (Society of Exploration Geophysicists, Tulsa, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The 3D seismic reflection data were acquired onboard RV *Maurice Ewing*, processed by Robertson Research Ltd, and interpreted using the 3D VoxelGeo software of Paradigm Geophysical. The ARAD seismic experiment was an international collaborative project between investigators from the University of Cambridge and Scripps Institution of Oceanography and was funded by the UK Natural Environment Research Council, British Institutions Reflection Profiling Syndicate, and the United States National Science Foundation. This is a Department of Earth Sciences (Cambridge) contribution and an Institut de Physique du Globe de Paris contribution.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.C.S. (singh@ipgp.jussieu.fr).

Magma-maintained rift segmentation at continental rupture in the 2005 Afar dyking episode

Tim J. Wright^{1†}, Cindy Ebinger^{2†}, Juliet Biggs¹, Atalay Ayele³, Gezahegn Yirgu⁴, Derek Keir² & Anna Stork¹

Seafloor spreading centres show a regular along-axis segmentation thought to be produced by a segmented magma supply in the passively upwelling mantle^{1,2}. On the other hand, continental rifts are segmented by large offset normal faults, and many lack magmatism. It is unclear how, when and where the ubiquitous segmented melt zones are emplaced during the continental rupture process. Between 14 September and 4 October 2005, 163 earthquakes (magnitudes greater than 3.9) and a volcanic eruption occurred within the ~60-km-long Dabbahu magmatic segment of the Afar rift, a nascent seafloor spreading centre in stretched continental lithosphere^{3,4}. Here we present a three-dimensional deformation field for the Dabbahu rifting episode derived from satellite radar data, which shows that the entire segment ruptured, making it the largest to have occurred on land in the era of satellite geodesy. Simple elastic modelling shows that the magmatic segment opened by up to 8 m, yet seismic rupture can account for only 8 per cent of the observed deformation. Magma was injected along a dyke between depths of 2 and 9 km, corresponding to a total intrusion volume of ~2.5 km³. Much of the magma appears to have originated from shallow chambers beneath Dabbahu and Gabho volcanoes at the northern end of the segment, where an explosive fissural eruption occurred on 26 September 2005. Although comparable in magnitude to the ten year (1975–84) Krafla events in Iceland⁵, seismic data suggest that most of the Dabbahu dyke intrusion occurred in less than a week. Thus, magma intrusion via dyking, rather than segmented normal faulting, maintains and probably initiated the along-axis segmentation along this sector of the Nubia–Arabia plate boundary.

Rifting of Africa and Arabia during the past ~30 Myr produced the ~300-km-wide Afar depression, which comprises the Afar triple junction. The three rift arms formed within a Palaeogene flood basalt province associated with the Afar mantle plume^{4,6,7}. The Dabbahu rift event occurred in the northern, or Red Sea, arm of the Afar triple junction (Fig. 1). Since ~4 Myr ago, faulting and volcanism within the northern Afar depression have localized to ~60-km-long axial volcanic ranges with aligned chains of basaltic cones, shallow seismicity and fissural flows, punctuated by stratovolcanoes. These volcanically and seismically active ‘magmatic segments’—Dabbahu (Boina), Alayta, Tat ‘Ale and Erta ‘Ale—are similar in size, morphology, structure and spacing to slow-spreading mid-ocean-ridge segments⁴. The highly extended and intruded crust beneath the Afar depression is ~18–22 km thick, with strong similarities to crust beneath Iceland^{6,8}. Global plate reconstructions based on geological and geodetic data suggest an average spreading rate of ~16 mm yr⁻¹ (ref. 9).

The current volcano-seismic crisis in the Dabbahu magmatic segment began on 14 September 2005 with an earthquake of body-wave magnitude (M_b) ~4.7 (Fig. 1; Supplementary Table SM1).

Nearly continuous seismic activity, including tremor, was registered at station FURI (Addis Ababa) between 24 and 26 September. Earthquake locations show that before 25 September, most seismicity occurred in the northern half of the segment, with the greatest earthquake density around Dabbahu and Gabho volcanoes (Fig. 1). The majority of earthquakes from 25 to 29 September occurred south of the two volcanoes. On 26 September, a 500-m-long, 60-m-deep, north–south oriented vent and faults opened on the eastern flank of Dabbahu volcano, south of a small volcanic cone known as Da’Ure (Supplementary Figs SM2–SM5). Initial analysis of pumice erupted

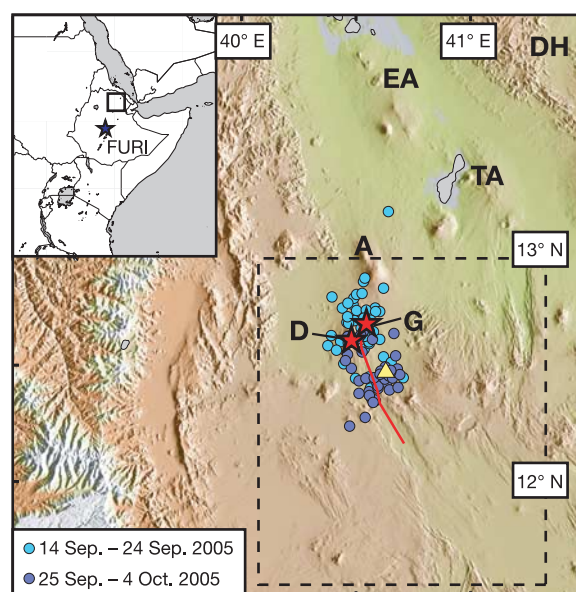


Figure 1 | Coloured and shaded relief map for northern Afar, and study area. Main figure shows the location of the dyke (red line) intruded along the entire Dabbahu magmatic segment, and Gabho (G) and Dabbahu (D) stratovolcanoes (red stars), superimposed on a coloured and shaded relief map derived from the Shuttle Radar Topographic Mission elevation model. Filled circles are earthquake epicentral locations (relative to the event marked by the yellow triangle) determined using the JED method, grouped before and after 24 September 2005. Most events occurred between 20 and 29 September 2005. The ~60-km-long Dabbahu, Alayta (A), Tat ‘Ale (TA) and Erta ‘Ale (EA) tectono-magmatic segments are the locus of Quaternary strain^{3,4}. To the northeast, the Afar depression is bounded by the northwest–southeast-trending Danakil horst (DH), and to the west, by the western escarpment. Dashed box encloses area shown in Fig. 2. Inset, location of study area with respect to East Africa and seismic station FURI, near Addis Ababa (star).

¹Department of Earth Sciences, University of Oxford, Oxford OX1 3PR, UK. ²Department of Geology, Royal Holloway, University of London, Egham TW20 0EX, UK.

³The Geophysical Observatory, ⁴Department of Earth Sciences, Addis Ababa University, Addis Ababa, Ethiopia. [†]Present addresses: School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK (T.J.W.); Department of Earth and Environmental Sciences, University of Rochester, Rochester, New York 14627, USA (C.E.).

from the vent is consistent with a felsic source at a depth of <6 km reheated by injection of basaltic magma.

We use radar data acquired by the European Space Agency's Envisat satellite to form ascending and descending interferograms that record the deformation in the Dabbahu event (Fig. 2a, b; Supplementary Table SM2). In addition we formed range and azimuth offset maps by correlating the radar amplitude images¹⁰ (Fig. 2c–f). Although these data are noisier than interferograms (Supplementary Table SM3), they do not require phase unwrapping and can provide measurements in places where interference fringes break down owing to high deformation gradients. By inverting these four scalar projections of the surface displacement field, where they overlap and are coherent, we have been able to estimate the three-dimensional displacement field of the Dabbahu rifting event¹¹ (Fig. 3a; Supplementary Methods).

The horizontal component of the deformation field at the surface shows a maximum opening of 6 m perpendicular to the Dabbahu rift segment (Fig. 3a). Along most of the rift, horizontal displacement vectors are fairly symmetrical and perpendicular to the centre of the rift. An approximately 25-km-wide zone centred on the rift was uplifted along most of the 60-km-long rift segment; uplift reached a maximum of ~ 1.5 m on both rift flanks (Fig. 3a, d). A 2–3-km-wide central zone, flanked by sharp discontinuities on either side, subsided by up to 2 m. At the northern end of the rift, 2–3 m of subsidence occurred around the Dabbahu and Gabho magmatic centres. Horizontal displacement vectors near the two volcanoes at the northern tip of the Dabbahu segment point radially inward.

Both the narrow zone of subsidence within the central rift and the shoulder uplift along the length of the Dabbahu segment are consistent with the displacement field expected for a large-scale vertical dyke intrusion with induced faulting^{5,12,13}. The dyke did not reach the surface along most of its length, suggesting that normal faulting occurred ahead of and/or above the laterally propagating dyke, enhancing the graben subsidence caused by the magma emplacement alone¹². The trends of the inferred normal faults and dyke closely match the orientations of faults previously mapped along the length of the Dabbahu segment⁴, and those found in the field (Supplementary Figs SM1–SM3).

To estimate the vertical distribution of opening, the volume of intruded magma, the amount of fault slip and the amount of magma sourced from the shallow chambers, we set up a simple elastic model that contains simplified geometries for all of the causative elements. The model has magma chambers beneath Dabbahu and Gabho volcanoes, represented by two point deflation sources¹⁴ at depths of 5 km, as suggested by initial analyses of erupted pumice. The dyke is modelled as a tensile dislocation in an elastic half space extending vertically from the surface to a depth of 10 km, following the centre of the observed subsidence, discretized into a series of rectangular patches, ~ 2 km long and 1 km wide¹⁵. Tests where we allowed the dyke to extend deeper showed that no opening was required outside this depth range. The near-surface normal faults are modelled as elastic dislocations dipping at 65° along the boundaries of the subsiding areas. They extend until they intersect the dyke—at 1 km depth in the north, 2 km in the centre and 2.5 km at the southern end, where the dyke is widest. The faults are discretized into 2-km patches along strike, and only normal slip is allowed.

Having fixed the model geometry, we carried out a simple least-squares inversion to determine the amount of dyke opening, fault slip and volume changes at the magma chambers that best fit the radar data, which were subsampled and weighted using the inverse of their full covariance matrices¹⁶. To prevent oscillatory solutions, the inversion was regularized with a positivity constraint and by requiring the spatial distribution to be smooth, the weight of the laplacian smoothing term being chosen as the best compromise between solution roughness and misfit. Uncertainties in our parameters were determined by Monte Carlo simulation of correlated noise¹⁷.

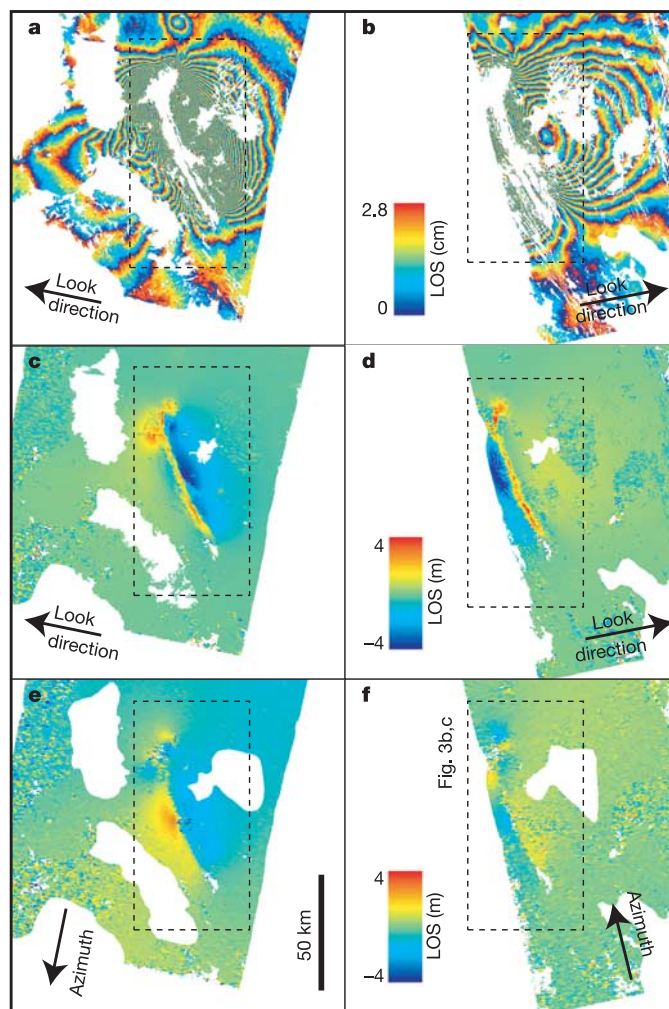


Figure 2 | Satellite radar data spanning the 2005 Dabbahu rifting event produced using data from ESA's Envisat satellite. Left column (a, c, e) from descending track 49 using images acquired on 6 May and 28 October 2005, IS2 beam mode; right column (b, d, f) from ascending track 29 images on 28 July 2005 and 26 October 2005. **a, b**, Wrapped interferograms; **c, d**, range change (unwrapped interferogram where available, range offsets elsewhere); **e, f**, azimuth offsets. LOS, satellite line of sight. Dashed boxes enclose area shown in Fig. 3. All SAR data copyright ESA.

Our best fitting model (Fig. 3b, c), which explains 92% of the data variance, suggests that 0.3 km^3 and 0.2 km^3 of contraction occurred in magma chambers beneath Dabbahu and Gabho, respectively. South of Dabbahu, the dyke opened by an average of 3.5 m, reaching a maximum of 8.0 m near the centre of the rift segment. The opening is greatest at 5–6 km depth, with little opening occurring below 9 km or above 2 km. Between the two magma chambers, the area coinciding with the greatest density of earthquakes, the modelled dyke opened by as much as 15 m. The modelled normal faults slipped by an average of 2 m, with as much as 7 m of slip inferred on the west of the rift segment above the area of maximum opening. Many discontinuities are evident in this area in the imagery, confirmed by helicopter-assisted field studies showing a ~ 2 -km-wide zone of open fissures, monoclines and normal faults, with up to 3 m of recent displacement (Supplementary Figs SM2–SM5).

Our analysis suggests that a total of $2.4\text{--}2.6 \text{ km}^3$ of magma was intruded along the Dabbahu rift segment during the ~ 2 -week rifting episode in September 2005. This single episode was twice as voluminous as estimates from the 1975–84 Krafla (Iceland) rifting events, when 9 m of opening was achieved through 20 dyking events and 9 eruptions⁵.

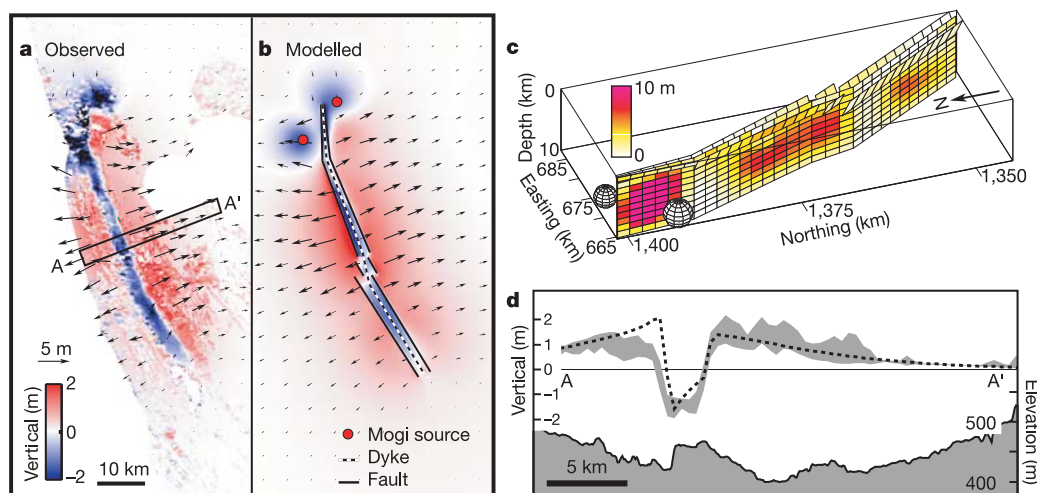


Figure 3 | Observed and modelled three-dimensional deformation field of the 2005 Dabbahu rifting episode. **a**, Three-dimensional deformation field created from interferograms and range and azimuth offsets (Supplementary Methods). Arrows represent the horizontal component and colours represent the vertical component. **b**, Modelled three-dimensional displacement field. **c**, Model using simplified geometry consisting of two deflating Mogi sources beneath Gabho and Dabbahu volcanoes, shown as

spheres, a vertical tensile dislocation with variable opening to represent the dyke and near-surface, dipping, normal faults. Colours indicated the amount of opening (dyke elements) or slip (fault elements). **d**, Profile across the dyke, showing (top) observed and modelled vertical deformation and (bottom) topography in the area outlined by the black rectangle in **a**. Grey band shows estimated 1σ errors for the vertical deformation field, and the black dashed line is from the model.

The apparent southward migration of seismicity is consistent with magma depletion from chambers beneath Gabho and Dabbahu volcanoes. However, this simple model suggests that contraction of the magma chambers can only account for 20% of the magma required to fill the dykes. It is possible that magma was intruded into the shallow magma chambers between the date of our last pre-rift radar acquisition (May 2005) and the rifting event (September 2005). Interferograms from November 2003 to April 2004 and April 2004 to May 2005 show no activity at Dabbahu for this entire period, but ~ 12 cm of uplift at Gabho volcano is observed between April 2004 and May 2005 (Supplementary Fig. SM6), consistent with ~ 0.01 km³ magma influx into that chamber. Unless the rate of deformation accelerated dramatically, it seems unlikely that a significant volume of magma was intruded between May and September 2005.

We also note that determining reliable volumes for magma removed from magma chambers using surface geodetic data alone is difficult^{18,19}. The results depend strongly on depth, Poisson's ratio (assumed to be 0.25 in these calculations) and the shape of the magma chamber. Furthermore, if the magma remaining in the chamber contains exsolved gases, it will expand when the pressure drops. The volume change of the magma chamber in these circumstances has previously been estimated to be as much as 4–5 times smaller than the volume of magma withdrawn from the chamber^{20,21}. We cannot therefore rule out the shallow chambers as being the source for all of the magma intruded into the dyke. Nevertheless, it is possible that additional magma has been sourced directly from deeper magma reservoirs. For example, a combination of magmatic inflation near the crust–mantle boundary and deflation of a shallow magma chamber were required to reproduce patterns in a 6-year series of interferograms in Iceland²². If, as in Iceland, the magma is distributed along the length of the rift segment, it would be difficult to extract such a signal from the much larger deformation caused by the shallow dyke intrusion. Further petrological, seismological and geodetic studies will allow us to better constrain the proportions of the two sources.

Seismic moment release during the period 14 September to 4 October 2005 offers additional insights into crustal deformation processes. We estimate the combined seismic moment release to be 6.7×10^{18} N m (Supplementary Methods), more than an order of magnitude smaller than our estimate for the geodetic moment

release ($\sim 8.0 \times 10^{19}$ N m). Hence, most of the deformation occurred aseismically. These short-term observations are also borne out by estimates of seismic efficiency over the past 40 years in Afar. Previous studies estimate that half the strain is accommodated aseismically²³, probably by episodic dyke intrusion^{24,25}. Structural data suggest that the locations of magma source(s) maintaining the Dabbahu segment have been stable for ~ 2 Myr (ref. 4).

Geodetic data corroborated with field and seismicity data clearly document deformation along the entire 60 km length of the Dabbahu rift segment, with the injection of a new column of crust up to 8 m wide. Without satellite radar interferometry, it would have been impossible to determine the full spatial extent of this event. Models of the surface deformation indicate that the source of at least 20% of the magma was shallow magma chambers near the northern tip of the segment, with any remainder directly emplaced from sources beneath the crust. The coincidence between the zone of magma intrusion and the previously mapped Dabbahu rift segment indicates that magmatic processes establish and maintain along-axis segmentation during continental rupture. Ongoing petrological, geochemical, seismic and geodetic studies will allow us to further probe the number and depth of magma sources, the temporal and spatial distribution of strain along and across the length of the Dabbahu magmatic segment, and the viscous relaxation^{26,27} of the lower crust and mantle beneath the rift.

Received 29 December 2005; accepted 8 June 2006.

- Lin, J., Purdy, G., Schouten, H., Sempère, J.-C. & Zervas, C. Evidence from gravity data for focused magmatic accretion along the Mid-Atlantic Ridge. *Nature* **344**, 627–632 (1990).
- Whitehead, J., Dick, J. H. B. & Schouten, H. A mechanism for magmatic accretion under spreading centres. *Nature* **312**, 146–148 (1984).
- Barberi, F. & Varet, J. Volcanism of Afar: Small-scale plate tectonic implications. *Geol. Soc. Am. Bull.* **88**, 1251–1266 (1977).
- Hayward, N. & Ebinger, C. Variations in the along-axis segmentation of the Afar rift system. *Tectonics* **15**, 244–257 (1996).
- Sigmundsson, F. *Iceland Geodynamics: Crustal Deformation and Divergent Plate Tectonics* (Springer-Praxis, Chichester, UK, 2006).
- Hirn, A., Lépine, J.-C. & Sapine, M. Triple junction and ridge hot spots: Earthquakes, faults, and volcanism in Afar, the Azores, and Iceland. *J. Geophys. Res.* **98**, 11995–12001 (1993).
- Ruegg, J.-C. *et al.* First epoch geodetic GPS measurements across the Afar plate boundary zone. *Geophys. Res. Lett.* **20**, 1899–1902 (1993).

8. Dugda, M. & Nyblade, A. in *The Afar Volcanic Province within the East African Rift System* (eds Yirgu, G., Ebinger, C. J. & Maguire, P. K. H.) 239–253 (Special Publication, Geological Society of London, 2006).
9. Kreemer, C., Holt, W. E. & Haines, A. J. An integrated global model of present-day plate motions and plate boundary deformation. *Geophys. J. Int.* **154**, 8–34 (2003).
10. Michel, R., Avouac, J. P. & Taboury, J. Measuring ground displacements from SAR amplitude images: application to the Landers earthquake. *Geophys. Res. Lett.* **26**(7), 875–878 (1999).
11. Wright, T. J., Parsons, B. E. & Lu, Z. Towards mapping surface deformation in three dimensions using InSAR. *Geophys. Res. Lett.* **31**, L01607, doi:10.1029/2003GL018827 (2004).
12. Rubin, A. & Pollard, D. Dike-induced faulting in rift zones in Iceland and Afar. *Geology* **16**, 413–417 (1988).
13. Abdallah, A. *et al.* Relevance of Afar seismicity and volcanism to the mechanics of accreting plate boundaries. *Nature* **282**, 17–23 (1979).
14. Mogi, K. Relations between the eruptions of various volcanoes and the deformation of the ground surfaces around them. *Bull. Earthq. Res. Inst.* **36**, 99–134 (1958).
15. Okada, Y. Surface deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **75**, 1135–1154 (1985).
16. Wright, T. J., Lu, Z. & Wicks, C. Constraining the slip distribution and fault geometry of the M_w 7.9, 3 November 2002, Denali fault earthquake with Interferometric Synthetic Aperture Radar and Global Positioning System data. *Bull. Seismol. Soc. Am.* **94**(6B), S175–S189 (2004).
17. Wright, T. J., Lu, Z. & Wicks, C. Source model for the M_w 6.7 23 October 2002 Nenana Mountain Earthquake (Alaska) from InSAR. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL018014 (2003).
18. McTigue, D. Elastic stress and deformation near a finite spherical magma body: resolution of the point source paradox. *J. Geophys. Res.* **92**(B12), 12931–12940 (1987).
19. Delaney, P. & McTigue, D. Volume of magma accumulation or withdrawal estimated from surface uplift or subsidence, with application to the 1960 collapse of Kilauea Volcano. *Bull. Volcanol.* **56**, 417–424 (1994).
20. Johnson, D., Sigmundsson, F. & Delaney, P. Comment on “Volume of magma accumulation or withdrawal estimated from surface uplift or subsidence, with application to the 1960 collapse of Kilauea Volcano” by P.T. Delaney and D.F. McTigue. *Bull. Volcanol.* **61**, 491–493 (2000).
21. Nishimura, T. Pressure recovery in magma due to bubble growth. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019810 (2004).
22. de Zeeuw-van Dalen, E., Pedersen, R., Sigmundsson, F. & Pagli, C. Satellite radar interferometry suggests deep accumulation of magma near the crust–mantle boundary beneath the Krafla volcanic system, Iceland. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020368 (2004).
23. Hofstetter, R. & Beyth, M. The Afar Depression: interpretation of the 1960–2000 earthquakes. *Geophys. J. Int.* **155**, 715–732 (2003).
24. Cattin, R. *et al.* Numerical modelling of Quaternary deformation and post-seismic displacement in the Asal-Ghoubbet rift (Djibouti, Africa). *Earth Planet. Sci. Lett.* **239**, 352–367 (2005).
25. Keir, D., Ebinger, C., Stuart, G., Daly, E. & Ayele, A. Strain accommodation by magmatism and faulting as rifting proceeds to breakup: Seismicity of the northern Ethiopian rift. *J. Geophys. Res.* **111**, B05314, doi:10.1029/2005JB003748 (2006).
26. Foulger, G. *et al.* Post-rifting stress relaxation at the divergent plate boundary in Northeast Iceland. *Nature* **358**, 488–490 (1992).
27. Pollitz, F. & Sacks, I. Viscosity structure beneath northeast Iceland. *J. Geophys. Res.* **101**(B8), 17771–17794 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Staff at Addis Ababa University, especially L. M. Asfaw, are thanked for their rapid response to the Dabbahu crisis, as are the Ethiopian Air Force for helicopter support. C. Oppenheimer, T. Kidane, A. Philpotts, D. Ayalew, G. Orsi and A. Asrat provided field reports and initial volcanological observations. This report benefited from discussions with B. Parsons, E. Calais, R. Buck, L. Asfaw, C. Wicks and J. Cann. Our work was supported by an NERC urgency grant, Addis Ababa University, the Ministry of Capacity Building of the Ethiopian Federal Government, and the Afar Regional Government. SAR data were provided quickly by the European Space Agency. T.J.W. is funded by a Royal Society University Research Fellowship.

Author Contributions T.J.W., C.E., G.Y. and A.A. planned the project; T.J.W. and J.B. processed, analysed and modelled the radar data; A.S., D.K. and A.A. analysed seismic data; G.Y. and C.E. provided petrological and tectonic context; and T.J.W., C.E., D.K. and J.B. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J.W. (t.wright@see.leeds.ac.uk).

An RNA-dependent RNA polymerase is required for paramutation in maize

Mary Alleman¹†, Lyudmila Sidorenko¹, Karen McGinnis¹, Vishwas Seshadri¹†, Jane E. Dorweiler¹†, Joshua White¹†, Kristin Sikkink¹ & Vicki L. Chandler¹

Paramutation is an allele-dependent transfer of epigenetic information, which results in the heritable silencing of one allele by another¹. Paramutation at the *b1* locus in maize is mediated by unique tandem repeats that communicate *in trans* to establish and maintain meiotically heritable transcriptional silencing². The *mop1* (*mediator of paramutation1*) gene is required for paramutation³, and *mop1* mutations reactivate silenced *Mutator* elements⁴. Plants carrying mutations in the *mop1* gene also stochastically exhibit pleiotropic developmental phenotypes³. Here we report the map-based cloning of *mop1*, an RNA-dependent RNA polymerase gene (RDRP), most similar to the RDRP in plants that is associated with the production of short interfering RNA (siRNA) targeting chromatin^{5,6}. Nuclear run-on assays reveal that the tandem repeats required for *b1* paramutation are transcribed from both strands, but siRNAs were not detected. We propose that the *mop1* RDRP is required to maintain a threshold level of repeat RNA, which functions *in trans* to establish and maintain the heritable chromatin states associated with paramutation.

To begin to elucidate the mechanism of paramutation, we pursued the cloning of *mop1* by using two alleles obtained in previous studies^{3,7}. A large backcross mapping population was created with *mop1-1* and the B73 inbred line (Supplementary Information) with the use of the phenotype of suppression of *B'* silencing³ to identify homozygous *mop1-1* plants (Supplementary Fig. 1).

Our initial mapping (Supplementary Tables 1 and 2) positioned the *mop1* gene to a region between the tightly linked *CL10221-1* marker and *umc1541*, 0.68 centimorgans away (Fig. 1a). We then exploited the high degree of synteny between the maize and rice genomes⁸ to identify rice sequences homologous to the maize mapping markers (Fig. 1b). In rice, the gene homologous to the maize *prp2* locus, which contains the tightly linked *phi083* and *CL10221-1* markers, rice *prp2* (Os04g39160)⁹, is adjacent to a gene encoding a RDRP (Fig. 1c). If a *rdp* gene was located near *prp2* in maize, it would be an excellent candidate for *mop1*, because this class of proteins functions in RNA-mediated interference (RNAi) pathways in multiple species (reviewed in refs 10, 11), including RNA-directed DNA methylation and the regulation of *trans*-acting siRNAs^{5,6}.

We located the tightly linked *prp2* gene on the maize physical map (<http://genome.arizona.edu/fpc/WebAGCoL/maize/WebFPC>) and with the use of polymerase chain reaction (PCR) analysis identified a *rdp* gene, *rdl101*, on the bacterial artificial chromosome (BAC) adjacent to the one carrying *prp2* (data not shown). These two BACs were sequenced and aligned to form a 261-kilobase-pair (kb) contig that contained five possible genes (Fig. 1d, and Supplementary Methods). Polymorphic markers were used to locate *mop1* more

precisely between the genes *stk2* and *prp2*, thus limiting candidate genes to *stk2*, *exp*, *rdl101* and *prp2* (Supplementary Table 1).

To test the hypothesis that *rdl101* was *mop1*, PCR and sequencing were used to screen for sequence changes that would disrupt gene function in the *rdl101* gene in two independent *mop1* mutations. The *mop1-1* allele was identified as a pre-existing mutation in a maize line containing active *Mutator* transposons³; *mop1-2* was isolated from an ethylmethane sulphonate (EMS) mutagenesis of maize pollen⁷. PCR reactions that overlapped a 300-base-pair (bp) region of the *rdl101* exon 4 in *mop1-1* plants failed to generate products, indicating a

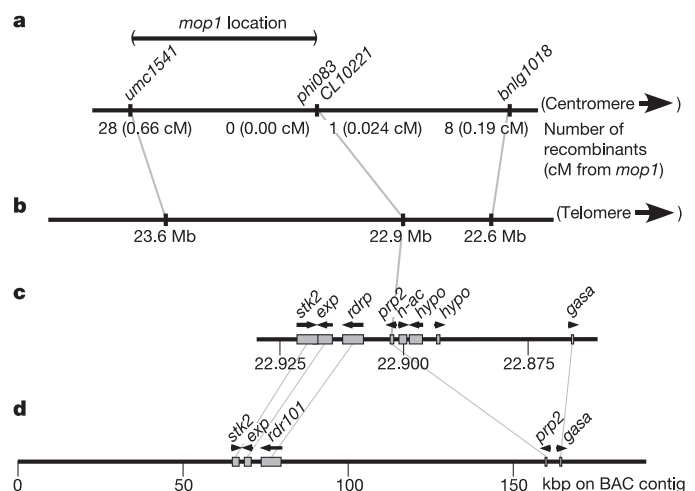


Figure 1 | Genetic mapping of *mop1* and identification of candidate genes. **a**, *mop1* was mapped to the indicated region of maize chromosome 2S with the use of PCR-based polymorphic markers (Supplementary Information). **b**, Diagram of the region of rice chromosome 4 (ref. 9) syntenic to the maize *mop1* region. Vertical grey lines (linking all pairs of diagrams) indicate the positions of homologous sequences. **c**, Gene content in the rice BAC AL606653 containing the homologue of the maize gene *prp2*, which is tightly linked to *mop1*. Genes or putative genes are shown as grey rectangles with arrows designating the presumed direction of transcription: *stk2* (*serine-threonine kinase-2*), *exp* (*expressed protein*), *rdp* (*RNA-dependent RNA polymerase*), *prp2* (*pathogenesis-related protein*), *n-ac* (*n-acetyltransferase*), *hypo* (*hypothetical protein*) and *gasa* (*gibberellic-acid-regulated protein*). **d**, Diagram showing the maize genes within the 261-kilobase-pair contig. No other non-transposon-related or non-retrotransposon-related genes were detected. The *rdl101* and *prp2* genes are separated by 81 kb of DNA that is 94% retrotransposon in origin.

¹Department of Plant Sciences, 303 Forbes Hall, University of Arizona, Tucson, Arizona 85721, USA. †Present addresses: Department of Biological Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, USA (M.A.); Biologics Development Center Developing Businesses, Dr Reddy's Laboratories Ltd, Survey no. 47, R. R. District, Andhra Pradesh 500072, India (V.S.); Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin 53201, USA (J.E.D.); Institute for Cellular and Molecular Biology, University of Texas at Austin, College of Natural Sciences, 1 University Station A4800, Austin, Texas 78712, USA (J.W.).

possible insertion within that region. Additional PCR reactions with *rdr101* primers and the TIR6 primer that amplifies all *Mutator* element termini (Supplementary Methods), followed by sequencing of the resulting amplifications from *mop1-1* plants, confirmed a *Mutator* element insertion in exon 4 of *rdr101* (Fig. 2a, b). PCR and sequencing revealed a single nucleotide G → A transition in *rdr101* in *mop1-2* relative to its progenitor, which causes a stop codon 37 amino-acid residues before the RDRP domain (Fig. 2a, c). G → A transitions account for most point mutations caused by EMS¹², which is consistent with *mop1-2* being an EMS-induced mutation. These experiments establish that *rdr101* is *mop1*.

There are multiple RDRP encoding genes in higher plants¹³ (www.chromdb.org). Multi-species alignment and phylogenetic analysis revealed that *mop1* is most similar to RDR2 in rice and *Arabidopsis* (Supplementary Fig. 2). *Arabidopsis* RDR2 is postulated to use siRNAs in a pathway that leads to the *de novo* DNA methylation of direct repeats⁶. The requirement of *mop1* for paramutation with its associated tandem repeats³ and silencing of *Mutator* elements⁴, each involving transcriptional silencing and chromatin changes, is consistent with a similar role for *mop1*. Although both *mop1* and RDR2 are involved in transcriptional silencing, mutations in the two genes differ with respect to developmental phenotypes. Plants containing *mop1* mutations are small and unhealthy, and flower late; some show developmental phenotypes that resemble epimutations in unlinked genes³, similar to those reported for mutations in *DDM1* (ref. 14), which encodes a chromatin remodelling protein^{15,16}. In contrast, no morphological or developmental phenotypes have been reported for RDR2 mutations in *Arabidopsis*. The differences could be because the RDR2 and MOP1 proteins have undergone functional diversification, because compensatory functions are provided by other RDRPs, or because there is differential transposable element reactivation in maize and *Arabidopsis*.

The identity of MOP1 as a RDRP strongly indicates that paramutation involves RNA-directed chromatin modification. The mechanism could be similar to that for centromeric heterochromatin formation in *S. pombe*¹⁷, which requires the RDRP activity of *Rdp1*, the sole RDRP in *S. pombe*, to perpetuate a self-enforcing RNAi loop involving siRNAs^{18,19}. Previous studies involving *b1* paramutation

have shown that unique, noncoding tandem repeats (seven copies of an 853-bp sequence) located 100 kb upstream are required for the *trans*-interactions mediating paramutation². The strength of paramutation is directly correlated with the number of repeats, and alleles that do not participate in paramutation have a single copy of this sequence². To address whether siRNAs might be associated with *b1* paramutation, as has been shown for *mop1*-regulated *Mutator* transposons²⁰, we searched for siRNA from the repeats in three genotypes, namely *b1-K55*, *B'* and *B-I*. The recessive *b1-K55* allele does not participate in paramutation and contains a single copy of the 853-bp sequence. The *B'* and *B-I* alleles, the silent and active forms of an identical sequence containing seven tandem repeats, have very different transcription levels of the coding region²¹. Paramutation occurs when *B'* and *B-I* homozygotes are crossed; the *B-I* allele is irreversibly changed to *B'*. A functional *mop1* gene is absolutely required for *b1* paramutation and to maintain *B'* transcriptional silencing³.

We were unable to detect repeat-specific siRNA (21–25 nucleotides) in any genotype (*B-I*, *B'* or *b*) despite using large quantities of enriched small RNA and employing sensitive hybridization conditions (Supplementary Information). The siRNAs may be extremely rare²² or undetectable, as suggested for the RNA-directed chromatin alteration at the *PAI* loci in *Arabidopsis*²³. They might be produced at specific developmental stages, or in specific tissues; *rdr2* mutants in *Arabidopsis* relieve transgene silencing in a subset of tissues²⁴. Alternatively, a distinct RNA-directed mechanism that does not use siRNAs might be involved as in *Caenorhabditis elegans*, in which the *EGO-1*-encoded RDRP, but not the Dicer or Drosha nucleases, is required for chromatin changes associated with meiotic silencing of unpaired chromosomal regions²⁵.

To further investigate the involvement of *trans*-acting RNA mechanisms in paramutation, we examined tandem repeat transcription by using nuclear run-on assays (Fig. 3). Initially the fragments indicated in Fig. 3a, along with other fragments further upstream and downstream (not shown), were tested with DNA probes that would detect transcripts on either strand. Only the regions shown in Fig. 3a produced signals above background (data not shown). Transcription was next examined with strand-specific RNA probes, revealing transcription on both strands (Fig. 3b). The differences in

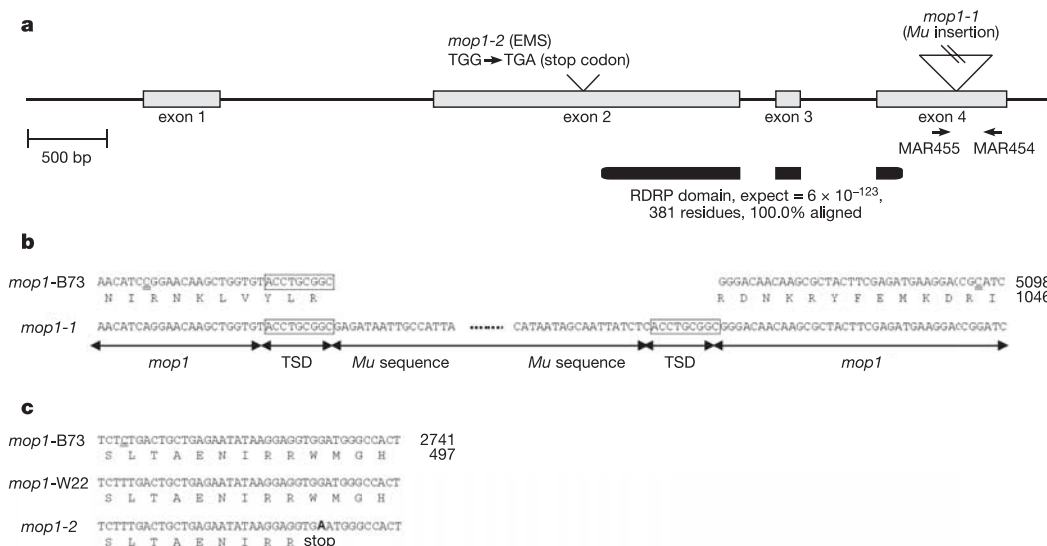


Figure 2 | The *mop1* gene is *rdr101*. Sequence locations are relative to the predicted start of translation (ATG). The double-underlined base pairs are polymorphisms in B73 relative to the progenitors for the mutant alleles. **a**, Diagram of *rdr101* showing the position of the *mop1-1* and *mop1-2* lesions, the RDRP domain and the primers used for PCR of the *Mutator* (*Mu*) insertion in *mop1-1*. The putative protein sequence of 1,127 amino

acids was based on splicing predictions *in silico* and multiple species alignments (Supplementary Methods). **b**, Sequence spanning the *mop1-1* insertion site shown in **a**. The target site duplication (TSD) begins at amino acid 1031. **c**, Sequence spanning the *mop1-2* EMS-induced sequence polymorphism; in comparison with B73 and the W22 progenitor, the nucleotide change replaces the codon at amino acid 494 with a stop codon.

the levels of transcription for each strand reflect relative differences in transcription rates because the opposite strands in each of these regions have very similar A/T percentages. In all three genotypes each region showed similar transcription levels (Fig. 3b) in spite of the fact that the *b1-K55* genotype has a single copy and *B'* and *B-I* have seven copies of each sequence. A summary of the results is shown diagrammatically in Fig. 3c.

We propose that the level of transcription occurring in the three genotypes is not sufficient to trigger paramutation or to maintain transcriptional silencing of *B'*, and that the RDRP encoded by *mop1* is required to produce a higher threshold of RNA (siRNAs or another type of RNA) that mediates these processes. Consistent with a threshold hypothesis is the observation that the penetrance of paramutation depends on the number of repeats². An intriguing question is why the transcribed single-copy sequence does not induce silencing. A hypothesis about the importance of tandem arrays in maintaining silencing in centromeric heterochromatin²⁶ offers a possible explanation (Fig. 4). Another idea is that the unique junction fragments created by the tandem repeats have specific properties. Because the penetrance of paramutation increases with the number of repeats², a mechanism that senses the number of repeat junctions needs to be postulated.

RNAi-directed heterochromatin formation at the centromeres in

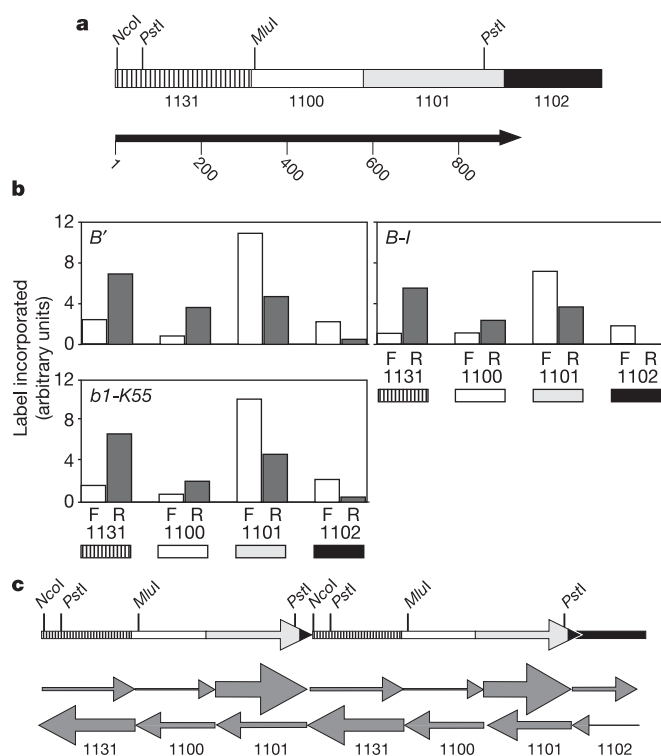


Figure 3 | Tandem repeats mediating *b1* paramutation are transcribed on both strands. **a**, Diagram of a single copy of the tandem repeats located 100 kb upstream of the *B-I* or *B'* start of transcription. The fragments of this repeat sequence used for the nuclear run-on assays are uniquely shaded and are designated as 1100, 1101, 1102 and 1131. **b**, Nuclei isolated from leaf sheaths of the three genotypes were subjected to nuclear run-on assays³. Isolated radioactive nuclear RNA was incubated with slot blots containing strand-specific RNA molecules for each of the sequences indicated in **a**, along with positive (*ubiquitin2*) and negative (synthetic RNA) controls. The histograms from one experiment illustrate the label incorporated into each indicated region normalized to *ubiquitin2*; F shows the amount of transcription from left to right, R from right to left. Another biological replicate gave comparable results (not shown). **c**, Diagram of two repeats (top) and a summary of nuclear run-on results. The sizes of the arrows (bottom) represent the relative transcription levels observed in the experiment shown in **b**.

*S. pombe*¹⁷ and *A. thaliana*²⁷ shares with *b1* paramutation the transcriptional control of associated genes, a requirement for RDRP, and tandem repeats. *RDR2* mutations rarely cause methylation changes within centromere repeats^{27,28}, in a similar manner to *mop1* mutations that do not cause demethylation of the centromere repeats or rDNA genes³. However, there are differences. To our knowledge there is no evidence that centromeres talk *in trans*; in contrast, the repeats associated with paramutation communicate *in trans* and this communication requires an RDRP. *RDR2* is required to establish methylation of the two tandem repeats at the *FWA* locus but is not required for methylation maintenance²², whereas *mop1* is required for both the establishment and maintenance of paramutation³. It remains to be determined whether these differences are a reflection of the repeats, functional overlap of a subset of activities by another RDRP, different types of RNA or different interpretations of RNA signals.

Our results add the classical paramutation phenomenon to a growing list of processes involving the RNA-mediated formation of chromatin domains. Some examples include dosage compensation in flies, X-inactivation in mammals, RNA-directed DNA methylation in plants, meiotic silencing of unpaired chromatin in *Neurospora* and *C. elegans*, and transcriptional silencing of centromeres, transgenes and transposons in multiple species^{11,25,29}. Like paramutation, most of these phenomena share epigenetic regulation, long-range interactions and chromosome dynamics. Unique features of *b1* paramutation are its extreme penetrance and faithful transmission through meiosis. Most of the other phenomena are either fully reset during meiosis and gamete formation or are only weakly heritable.

Paramutation models include *trans*-acting RNAs and direct interactions (pairing) between chromatin complexes associated with the tandem repeats¹. Our results provide strong support for the *trans*-acting RNA model¹, yet both *trans*-acting RNAs and pairing could be involved. Further studies will address whether both RNA and pairing are involved, how homologous sequences are sensed and counted, and how they communicate to form heritable chromatin structures.

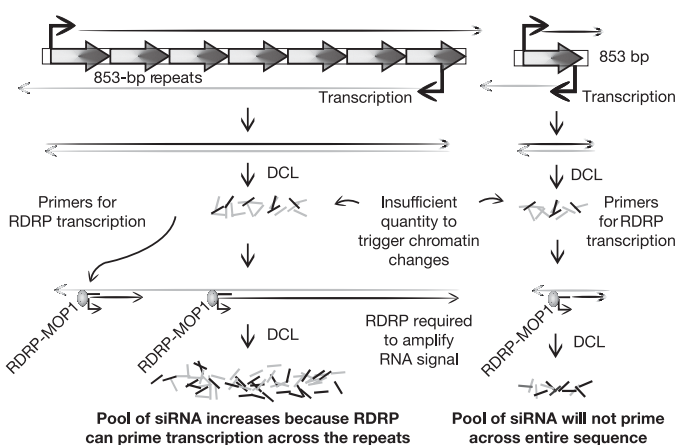


Figure 4 | Model for requirement of tandem repeats for RDRP amplification of RNA. The model shown is adapted from that in ref. 26. The involvement of RDRP in siRNA-induced silencing of transcribed tandem repeats (left) in comparison with unique or dispersed elements (right) is shown. Both the tandem repeats and the single-copy sequence are transcribed on both strands generating double-stranded RNA (dsRNA). Dicer-like (DCL) activity results in small quantities of siRNAs, which are not sufficient to trigger silencing. RDRP is required to increase the siRNAs above a threshold level to trigger silencing. Multiple rounds of RDRP and DCL with tandem repeats sustain increased pools of siRNA priming throughout the sequence (left). With a single-copy sequence, subsequent rounds of amplification would produce shorter and shorter dsRNAs (right). Although RDRPs do not necessarily require a primer³¹, increased amounts of primer may be required to efficiently reach the threshold level of RNA required for paramutation.

Although the mechanistic details of how RNA and chromatin interface are not known for any system, we suspect that both shared and distinct mechanisms are employed.

Note added in proof: A recent study has reported a paramutation-like phenomenon in the mouse, which is thought to involve an RNA-mediated post-transcriptional mechanism³⁰.

METHODS

Genetic stocks, mapping and cloning. For mapping, the *mop1-1* allele was crossed with the inbred line B73. Publicly available mapping markers were used to delimit the *mop1-1* region to an interval of chromosome 2 covered by 22 BACs. The two BACs containing and adjacent to the tightly linked markers were sequenced and the resulting contig was analysed to identify genes. Candidate genes were sequenced from the *mop1* mutants and progenitors. The recombination mapping results are summarized in Supplementary Tables 1 and 2. More detailed information on recombination mapping, BAC clone selection, sequencing and analysis, and candidate gene sequencing is given in Supplementary Methods.

Nuclear run-on analyses. Transcription was assayed with nuclear run-ons as described³. RNA sequences complementary to each strand were prepared by first amplifying the DNA template with either forward or reverse primers containing the T3 promoter as an overhang (Supplementary Methods). RNA was generated using the T3 Riboprobe system (Promega). Templates were digested with RNase-free DNase I and 20 ng of purified and concentrated RNA from each sample was applied to the slot blot in duplicate.

Received 3 March; accepted 11 May 2006.

- Chandler, V. L. & Stam, M. Chromatin conversations: mechanisms and implications of paramutation. *Nature Rev. Genet.* **5**, 532–544 (2004).
- Stam, M., Bebele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16**, 1906–1918 (2002).
- Dorweiler, J. E. *et al.* Mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12**, 2101–2118 (2000).
- Lisch, D., Carey, C. C., Dorweiler, J. E. & Chandler, V. L. A mutation that prevents paramutation in maize also reverses *Mutator* transposon methylation and silencing. *Proc. Natl Acad. Sci. USA* **99**, 6130–6135 (2002).
- Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
- Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336 (2004).
- Hollick, J. B. & Chandler, V. L. Genetic factors required to maintain repression of a paramutagenic maize *p1* allele. *Genetics* **157**, 369–378 (2001).
- Bennetzen, J. & Ma, J. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. *Curr. Opin. Plant Biol.* **6**, 128–133 (2003).
- Feng, Q. *et al.* Sequence and analysis of rice chromosome 4. *Nature* **420**, 316–320 (2002).
- Herr, A. J. Pathways through the small RNA world of plants. *FEBS Lett.* **579**, 5879–5888 (2005).
- Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**, 1635–1655 (2005).
- Kohalmi, S. E. & Kunz, B. A. Role of neighbouring bases and assessment of strand specificity in ethylmethanesulphonate and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine mutagenesis in the *SUP4-o* gene of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **204**, 561–568 (1988).
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543–553 (2000).
- Kakutani, T., Jeddeloh, J. A., Flowers, S. K., Munakata, K. & Richards, E. J. Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl Acad. Sci. USA* **93**, 12406–12411 (1996).
- Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genet.* **22**, 94–97 (1999).
- Brzeski, J. & Jerzmanowski, A. Deficient in DNA Methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *J. Biol. Chem.* **278**, 823–828 (2003).
- Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Sugiyama, T., Cam, H., Verdel, A., Moazed, D. & Grewal, S. I. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl Acad. Sci. USA* **102**, 152–157 (2005).
- Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
- Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nature Genet.* **37**, 641–644 (2005).
- Patterson, G. I., Thorpe, C. J. & Chandler, V. L. Paramutation, an allelic interaction, is associated with a stable and heritable reduction of transcription of the maize *b* regulatory gene. *Genetics* **135**, 881–894 (1993).
- Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
- Melquist, S. & Bender, J. Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in *Arabidopsis*. *Genes Dev.* **17**, 2036–2047 (2003).
- Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
- Maine, E. M. *et al.* EGO-1, a putative RNA-dependent RNA polymerase, is required for heterochromatin assembly on unpaired DNA during *C. elegans* meiosis. *Curr. Biol.* **15**, 1972–1978 (2005).
- Martienssen, R. A. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nature Genet.* **35**, 213–214 (2003).
- May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L. & Martienssen, R. A. Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. *PLoS Genet.* **1**, e79 (2005).
- Martienssen, R. A., Doerge, R. W. & Colot, V. Epigenomic mapping in *Arabidopsis* using tiling microarrays. *Chromosome Res.* **13**, 299–308 (2005).
- Lee, D. W., Seong, K. Y., Pratt, R. J., Baker, K. & Aramayo, R. Properties of unpaired DNA required for efficient silencing in *Neurospora crassa*. *Genetics* **167**, 131–150 (2004).
- Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
- Makeyev, E. V. & Bamford, D. H. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Wing and Arizona Genomics Institute scientists H. R. Kim, T. Rambo Mueller and F. Wei for sequencing and finishing the BAC clones; R. Jorgensen, C. Napoli and K. Gendler for performing the alignments and phylogenetic analyses; G. McCarthy and J. Bennetzen for help with the Bac-Breaker retroelement analysis of the BAC sequences; J. Gardiner for providing sequences for a mapping marker; and H. Basinger for technical assistance with the nuclear run-on assays. This work was supported by grants to V.L.C. from the National Science Foundation (NSF) and the National Institutes of Health; M.A. received a sabbatical supplement from the NSF. K.S., a student in the Undergraduate Biology Research Program, was supported in part by the Howard Hughes Medical Institute.

Author Contributions J.E.D. performed the initial mapping of *mop1*. L.S., J.W. and K.S. generated the large mapping population and performed the fine-structure mapping. M.A. compared the mapping data with the rice syntenic sequences, performed the sequencing, characterized the mutant alleles, and generated the figures. V.S. and K.M. performed the RNA and nuclear run-on experiments. V.L.C. directed the experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Sequences described in this paper have been deposited in GenBank under the following accession numbers: DQ417753 (ZMMBBb0178103), DQ417752 (ZMMBBb0004G18), DQ419917 (*mop1-1*), DQ417754 (*mop1-2*), DQ417755 (*rdi101* from W22) and DQ414253 (B73). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.L.C. (chandler@ag.arizona.edu).

Kruppel-like factor 2 regulates thymocyte and T-cell migration

Corey M. Carlson^{1*}, Bart T. Endrizzi^{1*}, Jinghai Wu², Xiaojie Ding¹, Michael A. Weinreich¹, Elizabeth R. Walsh¹, Maqsood A. Wani², Jerry B. Lingrel², Kristin A. Hogquist¹ & Stephen C. Jameson¹

Mammalian Kruppel-like transcription factors are implicated in regulating terminal differentiation of several tissue types^{1–3}. Deficiency in Kruppel-like factor (KLF) 2 (also known as LKLF) leads to a massive loss of the peripheral T-cell pool⁴, suggesting KLF2 regulates T-cell quiescence and survival^{4–7}. Here we show, however, that KLF2 is essential for T-cell trafficking. KLF2-deficient (*Klf2*^{−/−}) thymocytes show impaired expression of several receptors required for thymocyte emigration and peripheral trafficking, including the sphingosine-1-phosphate (S1P) receptor S1P₁, CD62L and β_7 integrin. Furthermore, KLF2 both binds and transactivates the promoter for S1P₁—a receptor that is critical for thymocyte egress and recirculation through peripheral lymphoid organs. Our findings suggest that KLF2 serves to license mature T cells for trafficking from the thymus and recirculation through secondary lymphoid tissues.

Kruppel-like factors (KLFs) are a family of zinc-finger transcription factors including at least 15 mammalian family members². KLFs have critical roles in the development of specific cell lineages, as demonstrated by the profound phenotypes that result from gene-targeting KLF family members^{1,2}. Several KLFs have an important role in cell maturation—a feature exemplified by erythroid KLF (KLF1), which is essential for erythrocyte production of adult-form haemoglobin^{1–3}.

KLF2 is expressed in lung, endothelial cells and lymphocytes^{4,8–11}, and is essential for normal blood-vessel integrity and lung development^{4,9–11}. Although *Klf2*^{−/−} thymocyte development is grossly normal, few KLF2-deficient T cells are found in peripheral lymphoid tissues^{4,6}. Moreover, the few *Klf2*^{−/−} T cells that are present in peripheral tissues show signs of activation and induction of cell death, suggesting KLF2 may have a critical role in T-cell quiescence and survival^{4,6}. KLF2 is normally expressed in mature thymocytes, naive T cells and memory T cells, but its expression is dramatically downregulated with T-cell receptor activation^{4,12–14}. Overexpression of KLF2 in the Jurkat T cell line leads to inhibition of cell-cycle progression, an effect that may involve KLF2 repression of *c-myc* transcription and/or induction of *p21*^{WAF1/CIP1} (refs 6, 15, 16). Taken together, these features have led to the proposal that KLF2 functions to prevent spontaneous activation and subsequent death of mature T cells^{4–7}.

To test this model, we studied maintenance of *Klf2*^{−/−} T cells *in vivo*. Because murine KLF2 deficiency is embryonic lethal^{9,11}, we generated fetal liver chimaeras (FLCs) using embryonic day 12.5 fetal livers from *Klf2*^{−/−} or *Klf2*^{+/−} donors injected into irradiated *Rag2*^{−/−} hosts (see Methods). After allowing haematopoietic reconstitution, we studied the T-cell pool in the thymus and peripheral lymphoid tissues of the chimaeras. Consistent with previous reports, which involved a distinct targeted allele of KLF2 (ref. 4), we found

that thymic development of *Klf2*^{−/−} T cells was grossly normal (Fig. 1a, b), but that there was a massive deficit of peripheral T cells (Fig. 1a, c). Notably, we also observed an increase in the representation of *Klf2*^{−/−} mature CD4⁺CD8[−] (CD4 single-positive; CD4 SP) and CD4[−]CD8⁺ (CD8 single-positive; CD8 SP) thymocytes compared with controls (Fig. 1b). Similar results were obtained using FLCs generated in lethally irradiated C57BL/6 (rather than *Rag2*^{−/−}) hosts (data not shown).

To study the proposed demise of mature *Klf2*^{−/−} T cells, we performed adoptive transfer of *Klf2*^{−/−} versus *Klf2*^{+/−} thymocytes into congenic hosts and tracked maintenance of the donor population. Unexpectedly, donor *Klf2*^{−/−} CD4 SP cells were recovered at similar total numbers compared to *Klf2*^{+/−} CD4 SP cells 14 days following adoptive transfer (Fig. 2a, b), and even at later time points (30 days post-transfer; data not shown)—results that seemed to conflict with the model that KLF2 was required for mature T-cell survival in the periphery. We considered that KLF2 deficiency might deregulate T-cell cytokine dependence, such that the normal survival cytokines were superfluous for *Klf2*^{−/−} T-cell persistence. However,

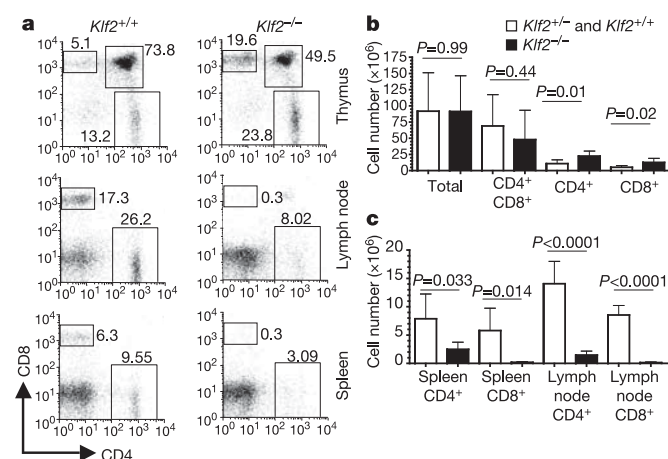


Figure 1 | *Klf2*^{−/−} T cells develop but do not populate the periphery in fetal liver chimaeras. Fetal liver chimaeras (FLCs) were generated from *Klf2*^{−/−} and *Klf2*^{+/−} (*Klf2*^{+/−} or *Klf2*^{+/+}) donors in *Rag2*^{−/−} hosts. **a**, Representative CD4/CD8 staining of the indicated tissues from *Klf2*^{−/−} and *Klf2*^{+/−} FLCs. Values represent the percentage of cells in each boxed region. **b**, **c**, Cell numbers for the indicated populations from thymus (**b**) and peripheral lymphoid tissues (**c**). Graphs show the average recovery from multiple experiments ($n > 6$ for **b**; $n > 4$ for **c**), with error bars representing standard deviation. Statistical comparisons between *Klf2*^{−/−} and *Klf2*^{+/−} groups are indicated.

¹Center for Immunology, Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA. ²Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati, Cincinnati, Ohio 45267, USA.

*These authors contributed equally to this work.

in vitro culture of $Klf2^{-/-}$ and $Klf2^{+/-}$ T cells showed very similar survival characteristics, including dependence on interleukin (IL)-7 for maintenance (Supplementary Fig. S1). As KLF2 deficiency has also been proposed to induce a loss in T-cell quiescence^{4,6}, it was possible that $Klf2^{-/-}$ T cells underwent spontaneous proliferation *in vivo*, balancing an increased rate of cell death. To explore this, we labelled proliferating cells *in vivo* with BrdU (5-bromodeoxyuridine). Both $Klf2^{-/-}$ and $Klf2^{+/-}$ T cells showed similar BrdU incorporation following adoptive transfer, arguing against differential proliferation of these populations (Fig. 2c). Although these data suggested $Klf2^{-/-}$ CD4 SP thymocytes were competent for short-term survival, we noted marked abnormalities in their tissue distribution: Whereas $Klf2^{+/-}$ donor T cells were found in blood, lymph nodes and spleen, $Klf2^{-/-}$ T cells segregated almost exclusively to the spleen (Fig. 2a, b), a pattern that was also seen at three and seven days following adoptive transfer (Supplementary Fig. S2a; data not shown). Recovery of both CD4⁺ and CD8⁺ donor $Klf2^{-/-}$ T cells was similar to controls soon after adoptive transfer (Supplementary Fig. S2a), but we did observe reduced recovery of $Klf2^{-/-}$ CD8⁺ (but not CD4⁺) T cells at later times (days 14 and 30; data not shown). However, it is currently unclear whether KLF2 has a direct role in long-term CD8⁺ T-cell survival, or if this gradual decline is secondary to altered T-cell trafficking.

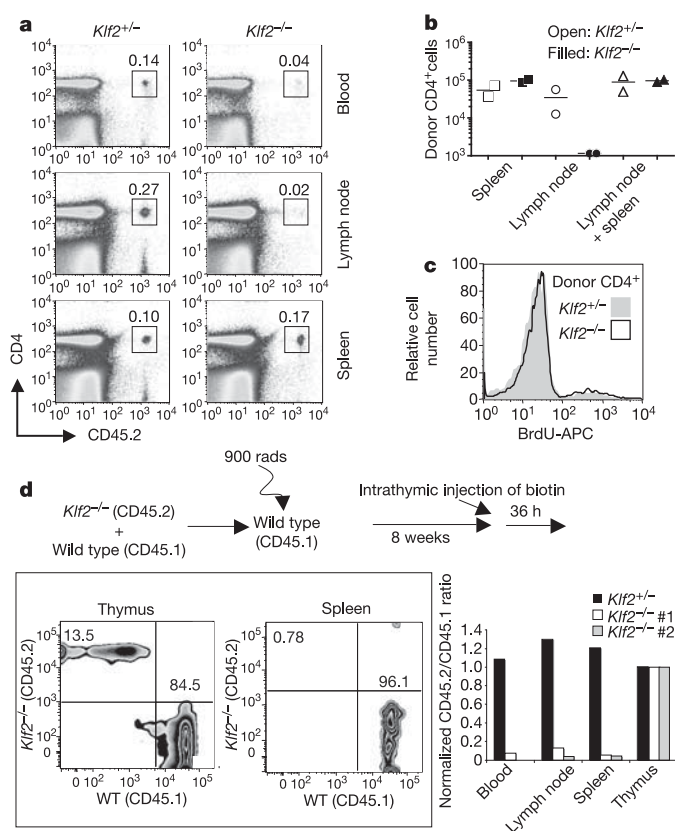


Figure 2 | $Klf2^{-/-}$ T cells survive but show deregulated trafficking following adoptive transfer. **a**, **b**, Thymocytes from $Klf2^{+/-}$ or $Klf2^{-/-}$ FLCs were transferred into C57BL/6.SJL hosts and labelled with BrdU for 14 days. The percentage (**a**) and absolute numbers (**b**) of donor CD4⁺ T cells in the indicated tissues were determined (representative of $n > 4$ experiments). **c**, Representative BrdU incorporation in donor-derived CD4⁺ splenocytes. An allophycocyanin (APC)-conjugated anti-BrdU antibody was used to detect BrdU in cells. **d**, The schematic describes the generation and intrathymic injection of mixed bone-marrow chimaeras. Fluorescence-activated cell sorting (FACS) plots show $Klf2^{-/-}$ and wild-type biotinylated CD4⁺ cells from a representative mixed chimaera. The bar graph indicates the ratio of FLC-derived:normal donor cells in indicated tissues, normalized against the thymic ratio.

These results suggested KLF2 regulates T-cell trafficking, raising the possibility that the absence of $Klf2^{-/-}$ peripheral T cells might arise from defective thymocyte emigration. To test this model directly, $Klf2^{-/-}$ or $Klf2^{+/-}$ FLCs were used to generate secondary radiation bone-marrow chimaeras, in which donor FLC cells were placed in competition with congenic wild-type cells (Fig. 2d). Biotin was then administered by intrathymic injection and, 36 h later, the export of biotin-labelled thymocytes into peripheral lymphoid tissues was determined. $Klf2^{+/-}$ recent thymic emigrants were detected in blood, spleen and lymph nodes in the ratios expected from thymocyte labelling (Fig. 2d). In contrast, $Klf2^{-/-}$ recent thymic emigrants were very rare, despite efficient labelling of these cells in the thymus (Fig. 2d). These data contrast with those expected if KLF2 deficiency caused T-cell death following thymic egress, in which case most peripheral $Klf2^{-/-}$ T cells would presumably be thymic emigrants. Hence, these results argue for impaired thymic emigration of $Klf2^{-/-}$ T cells.

To understand the basis for these altered trafficking patterns, we analysed the phenotype of $Klf2^{-/-}$ T cells. CD4⁺CD8⁺ (double-positive; DP) thymocytes from $Klf2^{-/-}$ and $Klf2^{+/-}$ chimaeras had very similar phenotypes (Fig. 3a), consistent with the fact that KLF2 is upregulated only after positive selection in the thymus^{4,17}. Single-positive (SP) thymocytes from $Klf2^{-/-}$ and $Klf2^{+/-}$ chimaeras showed similar expression of some markers, including T-cell receptor β (TCR β), CD25 and CD5 (Fig. 3a; data not shown), but other molecules were altered in KLF2-deficient cells. $Klf2^{-/-}$ SP thymocytes were CD69^{high} and CD62L^{low} (Fig. 3a), suggestive of a 'semi-mature' phenotype^{18–20}; yet these cells were also CD24^{low} and Qa2^{high} (Fig. 3a; data not shown), a phenotype consistent with full maturity^{18–20}. Despite their CD69 expression, the CD24^{low}CD25^{low} phenotype of $Klf2^{-/-}$ T cells argues against them being activated. Furthermore, expression levels of the activation/memory marker CD44 were similar on $Klf2^{-/-}$ and $Klf2^{+/-}$ thymocytes, although there was considerable variability in CD44 expression in individual FLCs (data not shown). With regard to trafficking, the expression of CD62L, CCR7 and β_7 integrin—all of which are involved in the entry of T cells into peripheral lymphoid tissues^{21–24}—was decreased on $Klf2^{-/-}$ T cells (Fig. 3a). Furthermore, CD69 expression by $Klf2^{-/-}$ SP thymocytes is potentially relevant, as CD69 can impair T-cell migration^{19,25,26}. The rare peripheral $Klf2^{-/-}$ T cells were phenotypically similar to SP thymocytes (Fig. 3a; data not shown), and the CD69^{hi}CD62L^{lo} phenotype of $Klf2^{-/-}$ SP thymocytes was maintained at 6 (Supplementary Fig. S2b), 14 or 30 (data not shown) days after adoptive transfer, indicating that the deregulated expression of these markers by $Klf2^{-/-}$ T cells is stable.

In order to further explore the maturation state and trafficking potential of $Klf2^{-/-}$ SP thymocytes, we sorted $Klf2^{-/-}$ and $Klf2^{+/-}$ DP and CD4 SP thymocytes and performed real-time polymerase chain reaction following reverse transcription (RT-PCR) on genes known to change in expression during thymocyte maturation (Supplementary Fig. S3). Most genes showed a similar pattern of expression in $Klf2^{+/-}$ and $Klf2^{-/-}$ thymocyte populations, consistent with normal thymocyte maturation in the absence of KLF2 (Supplementary Fig. S3). Real-time RT-PCR analysis did, however, demonstrate a dramatic decrease in the expression of mRNA encoding CD62L in $Klf2^{-/-}$ SP thymocytes (Fig. 3b and Supplementary Fig. S3) and a more modest impairment in the expression of β_7 integrin mRNA (Fig. 3b and Supplementary Fig. S3). Notably, loss of KLF2 had no substantial effect on the expression of CCR7 mRNA, although expression of the CCR7 protein was reproducibly reduced on $Klf2^{-/-}$ SP thymocytes (Fig. 3a). Also of note, two molecules involved in the regulation of apoptosis—Bcl2 and I α 1—were unaffected at the mRNA level by KLF2 loss (Fig. 3b and Supplementary Fig. S3). Although these assays focused on CD4 SP thymocytes, preliminary data indicate similar gene expression patterns in $Klf2^{-/-}$ CD8 SP cells (data not shown).

Reduced expression of CD62L, CCR7 and β_7 integrin immediately

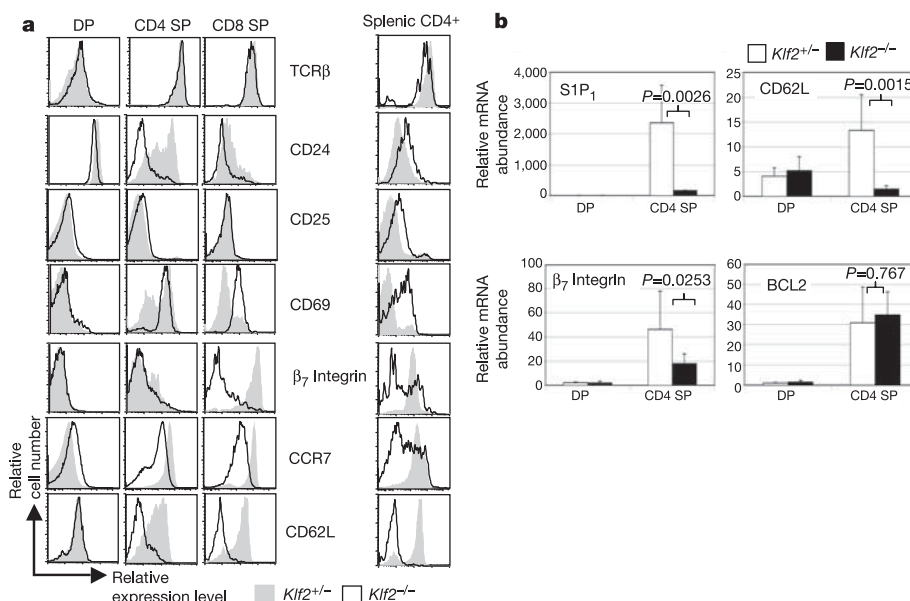


Figure 3 | KLF2 is required for thymocyte expression of critical trafficking molecules. **a**, Phenotype of thymocytes and CD4⁺ splenocytes from *Klf2*^{-/-} and *Klf2*^{+/+} FLCs (representative of $n > 4$ experiments). **b**, S1P₁, CD62L, β_7 integrin and BCL2 mRNA abundance was assessed by real-time

RT-PCR analysis of sorted DP and CD4 SP thymocytes from *Klf2*^{-/-} and *Klf2*^{+/+} FLCs. Bar graphs show the average signal from $n \geq 3$ experiments, with error bars representing standard deviation. *P*-values from comparisons between *Klf2*^{-/-} and *Klf2*^{+/+} CD4 SP samples are indicated.

suggested a basis for the altered trafficking of *Klf2*^{-/-} thymocytes: CD62L and CCR7 are essential for trafficking into peripheral lymph nodes; β_7 integrin for entry into Peyer's patches; and access to mesenteric lymph nodes requires either CD62L or β_7 integrin^{21–24}. However, although the loss of these molecules could explain aspects of altered *Klf2*^{-/-} T-cell trafficking in the adoptive transfer experiments (Fig. 2a, b), they were insufficient to account for the defect in *Klf2*^{-/-} thymocyte emigration (Fig. 2d). Recently, reports have demonstrated that the sphingosine-1-phosphate receptor S1P₁ (also known as Edg1) is critical for thymocyte emigration and T-cell recirculation^{26–28}. Hence, we examined the expression of S1P₁ in *Klf2*^{-/-} thymocytes. Real-time RT-PCR analysis demonstrated the expected upregulation of S1P₁ mRNA expression between the DP and CD4 SP stages in control thymocytes, but S1P₁ expression was not upregulated in *Klf2*^{-/-} CD4 SP thymocytes (Fig. 3b), nor in *Klf2*^{-/-} CD8 SP thymocytes (data not shown). On the basis of previous studies^{27–29}, defective S1P₁ expression could explain the crippled *Klf2*^{-/-} thymocyte egress observed (Fig. 2d), as well as the absence of *Klf2*^{-/-} T cells in peripheral blood following adoptive transfer (Fig. 2a, b).

These effects on gene transcription might indicate direct regulation by KLF2 or an indirect consequence of KLF2 deficiency altering thymocyte maturation. Given the primacy of S1P₁ in directing thymocyte egress, we focused on the ability of KLF2 to interact with the S1P₁ promoter. Chromatin immunoprecipitation (ChIP) assays were performed using an inducible KLF2 expression system^{6,15} as described previously¹⁴, and resulted in co-immunoprecipitation of KLF2 with the proximal S1P₁ promoter region (Fig. 4a). Furthermore, reporter assays demonstrated transactivation of the S1P₁ promoter by KLF2 (Fig. 4b), arguing that KLF2 positively regulates S1P₁ expression. These findings are consistent with previous reports showing upregulation of S1P₁ mRNA with KLF2 overexpression¹⁶. Interestingly, a consensus KLF family binding-site motif (CACCC) is found just upstream of the S1P₁ transcriptional start site (Supplementary Fig. S4), although KLF2 could also be acting through neighbouring SP1 sites, as demonstrated previously¹⁵. Together with our analysis of thymocyte gene expression, these findings suggest KLF2 directly induces S1P₁ expression in developing thymocytes.

In summary, our data suggest a radically different model of the role of KLF2: rather than primarily influencing T-cell survival, KLF2 regulates thymocyte and T-cell trafficking. Loss of KLF2 results in defective expression of S1P₁, CD62L, β_7 integrin and CCR7, and KLF2 seems to have a direct role in induction of S1P₁ expression. KLF2 and S1P₁ show similar expression patterns in T cells, both being upregulated on thymocyte maturation, downregulated after T-cell activation and re-expressed in the late effector/memory pool^{4,12,13,27}. Furthermore, KLF2-null and S1P₁-null mice die at similar stages in gestation because of widespread haemorrhaging, probably owing to defective tunica media integrity^{9,11,27,30}. It is interesting to speculate that KLF2 may be required for promoting S1P₁ expression in

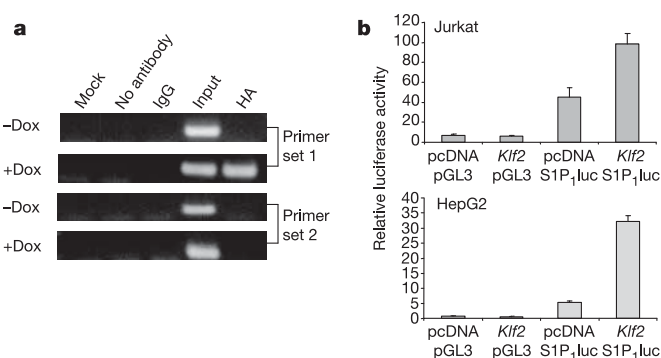


Figure 4 | KLF2 transactivates the S1P₁ promoter. **a**, Jurkat T cells were induced (“+Dox”) or not (“-Dox”) to express HA-tagged KLF2 and chromatin immunoprecipitation (ChIP) performed. PCR for S1P₁ promoter fragments was conducted on chromosomal DNA (“Input”), DNA co-precipitated with KLF2 (“HA”), and control immunoprecipitations, as indicated. Primer set 1 amplifies a region immediately upstream of the S1P₁ transcriptional start site (Supplementary Fig. S4), whereas primer set 2 amplifies a region about 300 bp further upstream. **b**, KLF2 transactivation of the S1P₁ promoter was assessed by reporter assay^{14,15}. Jurkat (T) and HepG2 (liver) cells were co-transfected with either *Klf2* cDNA (“*Klf2*”) or empty vector (“pcDNA”), plus either the S1P₁ promoter luciferase reporter (“S1P₁.luc”) or control plasmid (“pGL3”). Data are representative of $n = 3$ experiments. Error bars represent s.d.

endothelial cells, explaining the early embryonic lethality of KLF2-deficient mice.

Our data do not exclude additional roles for KLF2, potentially including long-term effects on T-cell quiescence or survival. Indeed, forced KLF2 expression can clearly promote cell-cycle withdrawal in tumour cell lines^{6,15,16}. However, our findings suggest KLF2 is critical to license mature thymocytes for trafficking competence. Such a role is consistent with the action of other KLF family members, which are essential for terminal differentiation of various cell types^{1–3}.

METHODS

A detailed description of all materials and methods can be found in Supplementary Information.

Mice, chimaeras and adoptive transfer. Fetal livers from embryonic day 12.5 *Klf2*^{−/−} and control animals were used to reconstitute irradiated *Rag2*^{−/−} hosts, which were analysed after eight weeks. In some experiments, thymocytes from FLCs were adoptively transferred into C57BL/6.SJL hosts, which were then offered BrdU-laced drinking water. Donor-derived (CD45.2⁺) cells were analysed for tissue distribution, phenotype and BrdU incorporation by flow cytometry, using the indicated antibodies (BD Pharmingen). Secondary bone-marrow chimaeras were generated using a mixture of bone marrow from FLCs (CD45.2) plus C57BL/6.SJL (CD45.1) animals to reconstitute lethally irradiated C57BL/6.SJL hosts. After eight weeks, these animals were intrathymically injected with biotin and the appearance of FLC-donor-derived (CD45.2) and wild-type-donor-derived (CD45.1) biotin-labelled cells in the thymus and periphery monitored.

Real-time RT-PCR. Thymocyte subsets were purified by fluorescence-activated cell sorting (FACS) on a FACSVantage (Becton Dickinson), followed by RNA isolation and preparation of cDNA. Real-time RT-PCR was performed on a SmartCycler (Cepheid) using primers listed in Supplementary Information. mRNA abundance was determined relative to controls (*Hprt*, *Gapdh* and/or *Cttnb1*).

Chromatin immunoprecipitation and gene reporter assays. Jurkat T cells transfected with tetracycline (tet)-inducible haemagglutinin (HA)-tagged KLF2 (refs 14, 15) were cultured with or without doxycycline for 48 h, formaldehyde-fixed, lysed, and subjected to immunoprecipitation with an anti-HA antibody or control IgG. Co-precipitated DNA was used as template for PCR of the S1P₁ promoter using primer sets 1 or 2 (initiating 141 base pairs (bp) and 474 bp upstream of the indicated S1P₁ transcriptional start site, respectively). For S1P₁ reporter assays, normal Jurkat or HepG2 cells were transiently transfected with the indicated plasmids and luciferase activity monitored at 48 h using a Monolight 3010 Luminometer (BD Biosciences).

Statistical analysis. An unpaired two-tailed Student's *t*-test was applied using GraphPad software (Prism) on normal or log₁₀-transformed data sets.

Received 18 December 2005; accepted 3 May 2006.

- Dang, D. T., Pevsner, J. & Yang, V. W. The biology of the mammalian Kruppel-like family of transcription factors. *Int. J. Biochem. Cell Biol.* **32**, 1103–1121 (2000).
- Kaczynski, J., Cook, T. & Urrutia, R. Sp1- and Kruppel-like transcription factors. *Genome Biol.* **4**, 206 (2003).
- Turner, J. & Crossley, M. Mammalian Kruppel-like transcription factors: more than just a pretty finger. *Trends Biochem. Sci.* **24**, 236–240 (1999).
- Kuo, C. T., Veselits, M. L. & Leiden, J. M. KLF2: a transcriptional regulator of single-positive T cell quiescence and survival. *Science* **277**, 1986–1990 (1997).
- Kuo, C. T. & Leiden, J. M. Transcriptional regulation of T lymphocyte development and function. *Annu. Rev. Immunol.* **17**, 149–187 (1999).
- Buckley, A. F., Kuo, C. T. & Leiden, J. M. Transcription factor KLF2 is sufficient to program T cell quiescence via a c-Myc-dependent pathway. *Nature Immunol.* **2**, 698–704 (2001).
- Di Santo, J. P. Lung Kruppel-like factor: a quintessential player in T cell quiescence. *Nature Immunol.* **2**, 667–668 (2001).
- Anderson, K. P., Kern, C. B., Crable, S. C. & Lingrel, J. B. Isolation of a gene encoding a functional zinc finger protein homologous to erythroid Kruppel-like factor: identification of a new multigene family. *Mol. Cell. Biol.* **15**, 5957–5965 (1995).
- Wani, M. A., Means, R. T. Jr & Lingrel, J. B. Loss of KLF2 function results in embryonic lethality in mice. *Transgenic Res.* **7**, 229–238 (1998).
- Wani, M. A., Wert, S. E. & Lingrel, J. B. Lung Kruppel-like factor, a zinc finger transcription factor, is essential for normal lung development. *J. Biol. Chem.* **274**, 21180–21185 (1999).
- Kuo, C. T. et al. The KLF2 transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. *Genes Dev.* **11**, 2996–3006 (1997).
- Schober, S. L. et al. Expression of the transcription factor lung Kruppel-like factor is regulated by cytokines and correlates with survival of memory T cells *in vitro* and *in vivo*. *J. Immunol.* **163**, 3662–3667 (1999).
- Grayson, J. M., Murali-Krishna, K., Altman, J. D. & Ahmed, R. Gene expression in antigen-specific CD8⁺ T cells during viral infection. *J. Immunol.* **166**, 795–799 (2001).
- Wu, J. & Lingrel, J. B. Kruppel-like factor 2, a novel immediate-early transcriptional factor, regulates IL-2 expression in T lymphocyte activation. *J. Immunol.* **175**, 3060–3066 (2005).
- Wu, J. & Lingrel, J. B. KLF2 inhibits Jurkat T leukemia cell growth via upregulation of cyclin-dependent kinase inhibitor p21^{WAF1/CIP1}. *Oncogene* **23**, 8088–8096 (2004).
- Haaland, R. E., Yu, W. & Rice, A. P. Identification of KLF2-regulated genes in quiescent CD4⁺ T lymphocytes. *Mol. Immunol.* **42**, 627–641 (2005).
- Mick, V. E., Starr, T. K., McCaughy, T. M., McNeil, L. K. & Hogquist, K. A. The regulated expression of a diverse set of genes during thymocyte positive selection *in vivo*. *J. Immunol.* **173**, 5434–5444 (2004).
- Gabor, M. J., Godfrey, D. I. & Scollay, R. Recent thymic emigrants are distinct from most medullary thymocytes. *Eur. J. Immunol.* **27**, 2010–2015 (1997).
- Feng, C. et al. A potential role for CD69 in thymocyte emigration. *Int. Immunol.* **14**, 535–544 (2002).
- Boursalian, T. E., Golob, J., Soper, D. M., Cooper, C. J. & Fink, P. J. Continued maturation of thymic emigrants in the periphery. *Nature Immunol.* **5**, 418–425 (2004).
- Arbones, M. L. et al. Lymphocyte homing and leukocyte rolling and migration are impaired in L-selectin-deficient mice. *Immunity* **1**, 247–260 (1994).
- Wagner, N. et al. Critical role for β7 integrins in formation of the gut-associated lymphoid tissue. *Nature* **382**, 366–370 (1996).
- Wagner, N. et al. L-selectin and β7 integrin synergistically mediate lymphocyte migration to mesenteric lymph nodes. *Eur. J. Immunol.* **28**, 3832–3839 (1998).
- Cyster, J. G. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu. Rev. Immunol.* **23**, 127–159 (2004).
- Nakayama, T. et al. The generation of mature, single-positive thymocytes *in vivo* is dysregulated by CD69 blockade or overexpression. *J. Immunol.* **168**, 87–94 (2002).
- Shiow, L. R. et al. CD69 acts downstream of interferon-α/β to inhibit S1P₁ and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544 (2006).
- Matloubian, M. et al. Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355–360 (2004).
- Allende, M. L., Dreier, J. L., Mandala, S. & Proia, R. L. Expression of the sphingosine-1-phosphate receptor, S1P₁, on T-cells controls thymic emigration. *J. Biol. Chem.* **279**, 15396–15401 (2004).
- Lo, C. G., Xu, Y., Proia, R. L. & Cyster, J. G. Cyclical modulation of sphingosine-1-phosphate receptor 1 surface expression during lymphocyte recirculation and relationship to lymphoid organ transit. *J. Exp. Med.* **201**, 291–301 (2005).
- Liu, Y. et al. Edg-1, the G protein-coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation. *J. Clin. Invest.* **106**, 951–961 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Cyster for generous provision of CCL19-Fc fusion protein, Y. Li for developing S1P₁ RT-PCR assays, M. Jenkins and M. Mescher for critical input on the manuscript, and members of the 'Jamequist' laboratory for helpful discussions. This work was supported by the NIH (an immunology pre-doctoral training grant to B.T.E., a grant to K.A.H., and a grant to S.C.J.), the American Cancer Society (a grant to S.C.J.), and the Cancer Research Institute (post-doctoral fellowship to C.M.C.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.C.J. (james024@umn.edu) or K.A.H. (hogqu001@umn.edu).

TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism

Xiaojing Pan¹, Sudharshan Eathiraj¹, Mary Munson¹ & David G. Lambright¹

Rab GTPases regulate membrane trafficking by cycling between inactive (GDP-bound) and active (GTP-bound) conformations¹. The duration of the active state is limited by GTPase-activating proteins (GAPs), which accelerate the slow intrinsic rate of GTP hydrolysis. Proteins containing TBC (Tre-2, Bub2 and Cdc16) domains are broadly conserved in eukaryotic organisms and function as GAPs for Rab GTPases as well as GTPases that control cytokinesis². An exposed arginine residue is a critical determinant of GAP activity *in vitro* and *in vivo*^{3–5}. It has been expected that the catalytic mechanism of TBC domains would parallel that of Ras and Rho family GAPs. Here we report crystallographic, mutational and functional analyses of complexes between Rab GTPases and the TBC domain of Gyp1p. In the crystal structure of a TBC-domain–Rab–GTPase–aluminium fluoride complex, which approximates the transition-state intermediate for GTP hydrolysis, the TBC domain supplies two catalytic residues *in trans*, an arginine finger analogous to Ras/Rho family GAPs and a glutamine finger that substitutes for the glutamine in the DxxGQ motif of the GTPase. The glutamine from the Rab GTPase does not stabilize the transition state as expected but instead interacts with the TBC domain. Strong conservation of both catalytic fingers indicates that most TBC-domain GAPs may accelerate GTP hydrolysis by a similar dual-finger mechanism.

Most known GAPs for Rab GTPases possess a TBC domain^{6–15}. Several yeast TBC-domain GAPs, termed Gyps (GAPs for Ypt/Rab proteins), have been shown to accelerate GTP hydrolysis for Ypt/Rab GTPases^{6–8,10,12,13}. One of the best-characterized TBC-domain GAPs, Gyp1p, has high catalytic activity for Ypt1p, Sec4p and other yeast

Rab GTPases^{3,7}. *In vivo*, Gyp1p localizes to the Golgi, negatively regulates Ypt1p and is required for the recycling of internalized cargo^{5,16}.

To identify substrates for co-crystallization, we investigated the ability of purified TBC domains to accelerate GTP hydrolysis for 29 mammalian Rab GTPases (Supplementary Fig. S1 and Supplementary Table S1). As shown in Fig. 1, the Gyp1p TBC domain has high selectivity for Rab1 (the mammalian homologue of Ypt1p) and Rab33, with catalytic efficiencies equivalent to the *in vivo* substrate Ypt1p. Like Ypt1p, Rab1 and Rab33 localize to the Golgi apparatus^{17–19}.

Aluminium fluoride (AlF_n) mimics the transition state for GTP hydrolysis^{20,21} and further stabilizes complexes of GDP-bound Ras and Rho family GTPases with cognate GAPs^{22–25}. When mixed in a 1:1 ratio and analysed by gel filtration, a significant fraction of GDP-bound Rab33 elutes with the Gyp1p TBC domain in the presence of AlF_n (Supplementary Fig. S2), indicating the formation of a stable ternary complex.

Diffraction-quality crystals were subsequently prepared for the Gyp1p TBC domain in complex with GDP-bound Rab33 and AlF_n. The structure was solved by multi-wavelength anomalous diffraction and refined to 2.3 Å (Supplementary Table S2 and Supplementary Fig. S3). As described previously⁴, the Gyp1p TBC domain contains 15 helices distributed between an amino-terminal subdomain composed of αG1–αG8 and a carboxy-terminal subdomain spanning αG9–αG15 (Fig. 2). Three signature motifs (RxxxW, IxxDxxR and YxQ) are located in well-ordered regions of the N-terminal subdomain. Residues from the IxxDxxR and YxQ motifs, which are

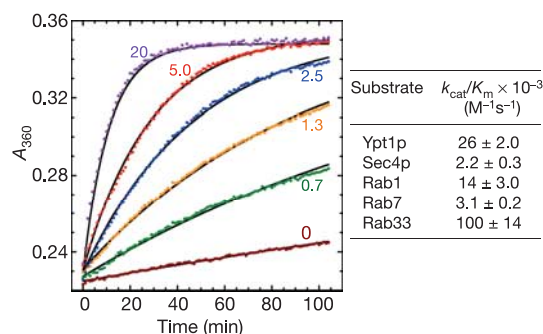


Figure 1 | Identification of mammalian Rab GTPase substrates for the Gyp1p TBC domain. Left: time courses of GTP hydrolysis for Rab33 in the absence and presence of the Gyp1p TBC domain (concentrations in nM are shown against the data). Solid lines represent a simultaneously fitted pseudo-first-order Michaelis–Menten model function (see Methods). Right: tabulation of catalytic efficiencies. Results are means $\pm$ s.d. for two independent experiments.

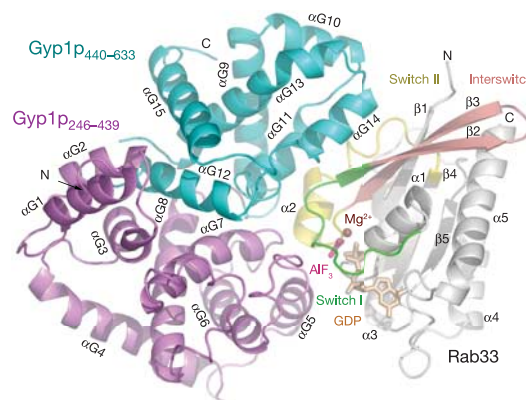


Figure 2 | Structure of the Gyp1p TBC domain in complex with Rab33–GDP–AlF₃. Shown is a ribbon representation of the complex with the N-terminal and C-terminal subdomains of Gyp1p and functional regions of Rab33 coloured as indicated by the labelling. No electron density was observed for the loop region between αG12 and αG13.

¹Program in Molecular Medicine & Department of Biochemistry and Molecular Pharmacology, UMASS Medical School, Two Biotech, 373 Plantation Street, Worcester, Massachusetts 01605, USA.

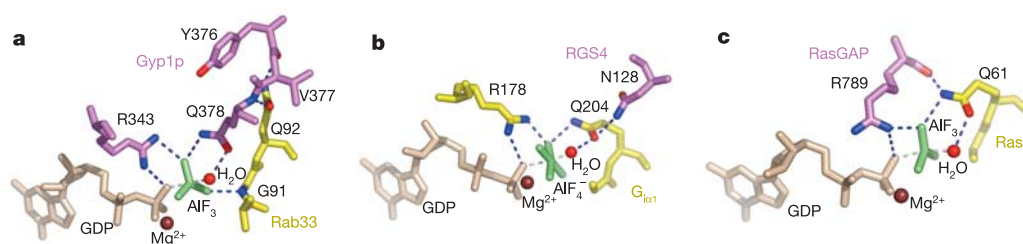


Figure 3 | Network of polar interactions in the AlF_3 -binding site and comparison with other GAP-GTPase complexes. a, Interactions in the AlF_3 -binding site in the Gyp1p-Rab33-GDP- AlF_3 complex. **b**, Interactions

in the AlF_3 binding site in the RGS4- $\text{G}\alpha_{11}$ -GDP- AlF_3 complex (Protein Data Bank (PDB) ID code 1AGR). **c**, Interactions in the AlF_3 -binding site in the RasGAP-Ras-GDP- AlF_3 complex (PDB ID code 1WQ1).

crucial to the hydrolytic mechanism as discussed below, are located in ordered loop regions at the C terminus of αG5 and at the N terminus of αG7 , respectively. Rab33 has a canonical GTPase fold with a central β -sheet ($\beta 1$ – $\beta 6$) surrounded by helices ($\alpha 1$ – $\alpha 5$).

Rab33 is cradled in a prominent rectangular groove formed by adjacent surfaces from the N-terminal and C-terminal subdomains of Gyp1p. The interaction surface in the N-terminal subdomain, which consists primarily of residues from the $\alpha\text{G5}/\alpha\text{G6}$ and $\alpha\text{G6}/\alpha\text{G7}$ loops, directly engages the GDP- AlF_n moiety and flanking elements in the switch regions and P-loop. The interaction surface in the C-terminal subdomain is formed by residues from the $\alpha\text{G8}/\alpha\text{G9}$ loop, αG11 and connecting loops, and αG14 . The latter surface interfaces with elements of the switch and interswitch regions distal to the GDP- AlF_n -binding site. The observed interaction epitope for Rab33 is consistent with previous studies, which implicated the rectangular groove as a likely binding site for Rab GTPases⁴. With the exception of three residues from the signature motifs, the Rab interaction epitope is poorly conserved (Supplementary Fig. S4).

In the experimental map, clear electron density is observed for Gyp1p, Rab33, GDP and Mg^{2+} in addition to a trigonal bipyramidal

arrangement of light atoms in the site normally occupied by the γ -phosphate in the GTP-bound state (Supplementary Fig. S3). The latter matches the expected geometry for AlF_3 , with a central Al^{3+} ion coordinating three equatorial fluoride ions, an oxygen from the β -phosphate of GDP and the putative nucleophilic water. Equivalent coordination spheres and trigonal bipyramidal geometries have been described for other GAP-GTPase complexes with AlF_3 (refs 23, 25).

Ras and Rho family GAPs supply an arginine residue, termed the 'arginine finger', which stabilizes the AlF_n complex through a polar/electrostatic interaction with one of the equatorial fluoride atoms^{23–25}. As predicted^{3–5}, the Gyp1p TBC domain provides an analogous arginine finger (Arg 343 in the IxxDxxR motif), which mediates hydrogen-bonding interactions with the axial oxygen from GDP and an equatorial fluoride ion (Fig. 3a). The stereochemistry resembles that observed for the *cis*-arginine finger in the $\text{G}\alpha$ subunits of G proteins^{20,21,26,27}.

In other GTPase- AlF_n complexes, a *cis*-glutamine residue from the DxxGQ motif mediates a bidentate interaction with an equatorial fluoride ion and the axial water^{20–27}. The stereochemistry is conserved between the Ras, Rho and G-protein families (Fig. 3b, c). With few

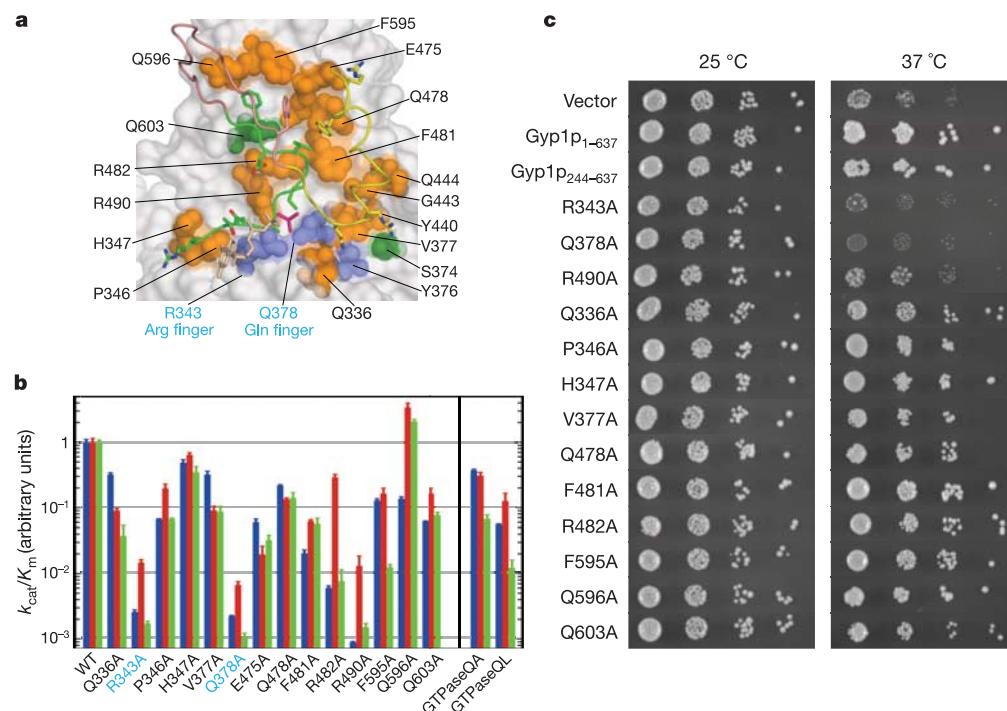


Figure 4 | Mutational analysis of the Gyp1p Rab interaction epitope. a, Location and conservation of residues in the Rab33 interaction epitope of Gyp1p. Blue, conserved; green, similar; orange, variable. **b**, Effect of mutations on the catalytic efficiency of Gyp1p. Results are means $\pm$ s.d. for two independent experiments. The mutant and wild-type proteins showed

comparable expression levels and solubilities. Blue, Rab33; red, Sec4p; green, Ypt1p. **c**, Plate assay testing the ability of wild-type and mutant Gyp1p constructs to rescue the temperature-sensitive slow-growth phenotype of a *GYPI*-null strain. Tenfold serial dilutions of normalized cell cultures were spotted from left to right.

exceptions, Rab GTPases also conserve the glutamine in the DxxGQ motif. Surprisingly, the corresponding glutamine in Rab33 does not contact the equatorial fluoride ions or axial water but instead mediates a bipartite polar interaction with the backbone carbonyl of Tyr 376 and amino group of Gln 378 in the YxQ motif of Gyp1p. The resulting conformation exposes the AlF_3 -binding site, allowing Gln 378 to mediate bipartite hydrogen-bonding interactions equivalent to the *cis*-glutamine in other GTPase- AlF_n complexes^{20–27}.

Given the close approximation of the $\text{GDP-AlF}_3\text{-H}_2\text{O}$ complex to the expected transition-state geometry for GTP hydrolysis, we propose that the catalytic mechanism proceeds through a structurally similar pentavalent intermediate stabilized by polar/electrostatic interactions with two distinct catalytic residues supplied *in trans*: the arginine finger from the IxxDxxR motif and a 'glutamine finger' from the YxQ motif. Broad conservation of both residues in all yeast TBC domains and at least 70% of mammalian TBC domains suggests that a dual-finger catalytic mechanism will be a general feature of the GTP hydrolytic reactions catalysed by most TBC domains. Atypical TBC domains (for example Tre17 isoforms), which lack one or usually both catalytic fingers, have no known GTPase substrates and either do not function as GAPs or use an alternative mechanism. Because all known TBC domain substrates conserve the glutamine in the DxxGQ motif, the bipartite interaction with backbone atoms in the YxQ motif is also likely to be a characteristic feature. Furthermore, some GAPs and cognate GTPases lack both an arginine finger and a *cis*-glutamine. In lieu of the canonical catalytic residues, Rap1GAP has been shown to supply a catalytic asparagine thumb²⁸, which was predicted to have a function equivalent to that observed here for the glutamine finger in the Gyp1p TBC domain.

Superposition of Gyp1p alone and in the complex with Rab33 reveals a small, roughly 2.5-Å, translation of the N-terminal subdomain relative to the C-terminal subdomain (Supplementary Fig. S5). Such flexibility between subdomains might help to facilitate the transition from the active conformation of the GTPase to the transition-state intermediate. Indeed, the switch regions of Rab33 undergo a significant conformational rearrangement in comparison with the GppNHp-bound form. Although we cannot exclude the possibility that the switch I and/or II regions are mobile in solution, it is likely that some if not all of the observed changes, particularly near the γ -phosphate/ AlF_3 -binding site, are driven by the interaction with Gyp1p. Finally, the substrates for which the Gyp1p TBC domain has the highest catalytic efficiency (Ypt1p, Sec4p and Rab33) have distinctly different active conformations²⁹, which are not complementary to the corresponding interaction epitopes in Gyp1p. Thus, significant structural changes would also be expected for Ypt1p and Sec4p.

Mutation of the arginine finger (Arg 343) or glutamine finger (Gln 378) to alanine results in a 100–1,000-fold decrease in catalytic efficiency (Fig. 4). Alanine substitutions involving other residues in the Rab interaction epitope decreases catalytic efficiency 2–500-fold. The large effects of several non-finger residues, notably Arg 490, probably reflect defects in binding and/or indirect effects on the stability of the active or transition state. In contrast, replacing the glutamine in the DxxGQ motif with alanine results in a relatively small 3–10-fold decrease in catalytic efficiency. Although replacement with leucine results in a larger 10–100-fold decrease, the effect is substantially less severe than that observed for the replacement of either catalytic finger with alanine.

High K_m values (more than 200 μM) for Ypt1p and Sec4p prevent a more detailed analysis. However, for Rab33, which has a lower K_m , the effect of the finger mutations is almost entirely on k_{cat} (Supplementary Fig. S6). Conversely, the decreased catalytic efficiency observed for the glutamine to leucine substitution in the DxxGQ motif is primarily the result of an order of magnitude increase in K_m without a significant decrease in k_{cat} . A similar pattern of large decreases in catalytic efficiency resulting from mutation of the arginine or glutamine fingers combined with relatively small effects

for replacement of the *cis*-glutamine is observed for Gyp7p and Ypt7p (Supplementary Fig. S6). Similarly, a much smaller than expected threefold decrease for the glutamine to leucine mutation in Ypt1p has previously been described for the Gyp5p-catalysed reaction, despite a 40-fold decrease in the intrinsic rate³⁰.

Budding yeasts lacking Gyp1p exhibit a temperature-sensitive growth defect on synthetic medium, which can be rescued by the expression of full-length Gyp1p or the isolated TBC domain (Gyp1p_{212–637}) from a *CEN* plasmid under control of the endogenous promoter⁵. In contrast, an arginine-finger mutant (R343K) fails to rescue even though it expresses at wild-type levels⁵. To establish whether the dual-catalytic-finger mechanism operates *in vivo*, we compared the ability of 13 Gyp1p mutants to rescue the *GYP1*-null phenotype (Fig. 4c). As expected, there were no discernable differences at the permissive temperature, which is consistent with the absence of secondary dominant-negative effects. At the non-permissive temperature, neither the glutamine-finger nor the arginine-finger mutants are capable of rescuing the null phenotype. The R490A substitution also fails to rescue, whereas other mutants exhibit partial or complete rescue of the growth defect. Overall, the ability of Gyp1p mutants to rescue is correlated with the catalytic efficiency measured *in vitro*. Consistent with this is the observation that the expressed protein levels for all 13 mutants are comparable to those of the wild-type protein (Supplementary Fig. S7). We therefore conclude that the failure of the glutamine/arginine finger and R490A mutants to restore normal growth reflects the loss of functional interactions with an *in vivo* Rab GTPase target (probably Ypt1p).

Mutation of the *cis*-glutamine to leucine is widely used to generate constitutively active GTPases. However, the effect on the rate of GTP hydrolysis catalysed by TBC domains is highly variable and in some cases much weaker than the effect on the intrinsic rate³⁰. Evidently, the *cis*-glutamine participates directly in transition-state stabilization for the intrinsic reaction but is co-opted by the *trans*-glutamine finger in the reaction catalysed by the TBC domain. In cases in which the *in vivo* rate of GTP hydrolysis is determined primarily by TBC-domain GAPs, the underlying mechanism for constitutive activation would involve diminished affinity for the GAP rather than elimination of a catalytic residue. Finally, TBC domains with mutated glutamine fingers should provide useful reagents for substrate identification as well as functional analyses *in vivo*.

METHODS

Constructs, expression and purification. Constructs and procedures for expression and purification are described in Supplementary Methods.

GAP assay. Rab GTPases were loaded with GTP by incubating 2–3 mg of protein with a 25-fold molar excess of GTP at 25 °C for 1 h in 20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM EDTA, 1 mM dithiothreitol. Free nucleotide was removed with a D-Salt column (Pierce Biotechnology) pre-equilibrated with 20 mM HEPES pH 7.5, 150 mM NaCl. The single-turnover kinetics of intrinsic and GAP-accelerated GTP hydrolysis were measured by a continuous enzyme-coupled optical assay for the release of inorganic phosphate with the use of reagents from the EnzChek Phosphate Assay Kit (Invitrogen). GTP-loaded Rab GTPases were mixed with solutions containing the assay reagents and GAPs, and dispensed into 96-well half-area microplates (Corning) with a Precision 2000 pipetting system (Bio-Tek). The final solutions contained 20 mM HEPES pH 7.5, 150 mM NaCl, 0.15 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside, 0.75 U ml^{-1} purine nucleoside phosphorylase, 10 mM MgCl_2 , 20 μM GTP-loaded Rab GTPases and various concentrations of GAPs. The absorbance at 360 nm was monitored with a Safire microplate spectrometer (Tecan). Data were analysed by fitting them simultaneously to the pseudo-first-order Michaelis-Menten model function

$$A(t) = (A_{\infty} - A_0)(1 - \exp(-k_{\text{obs}} t)) + A_0$$

where $k_{\text{obs}} = k_{\text{intr}} + (k_{\text{cat}}/K_m)[\text{GAP}]$. The catalytic efficiency (k_{cat}/K_m) and intrinsic rate constant for GTP hydrolysis (k_{intr}) were treated as global parameters.

Yeast plate assay. Yeast strains and plasmids are described in Supplementary Methods. The *gyp1Δ* cells (yeast strain NY2292) were grown at 25 °C overnight in Synthetic Complete (SC) medium (–URA) on a roller drum. *CEN LEU2* plasmids expressing different Gyp1p wild-type and mutant constructs were

transformed into *gyp1Δ* cells and plated on SC (–URA, –LEU) plates. Transformants were inoculated into SC (–URA, –LEU) medium and grown to exponential phase at 25 °C. The cultures were normalized to equivalent cell densities and diluted 1:10 in serial increments; equal volumes were spotted on SC (–URA, –LEU) plates. Images were acquired after 3 days of incubation at 25 or 37 °C.

Crystallization, data collection and structure determination. Procedures for crystallization, data collection and structure determination are described in Supplementary Methods.

Received 5 January; accepted 28 April 2006.

- Segev, N. Ypt/Rab GTPases: regulators of protein trafficking. *Sci. STKE* **2001**, RE11 (2001).
- Bernards, A. GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and *Drosophila*. *Biochim. Biophys. Acta* **1603**, 47–82 (2003).
- Albert, S., Will, E. & Gallwitz, D. Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J.* **18**, 5216–5225 (1999).
- Rak, A. *et al.* Crystal structure of the GAP domain of Gyp1p: first insights into interaction with Ypt/Rab proteins. *EMBO J.* **19**, 5105–5113 (2000).
- Du, L. L. & Novick, P. Yeast rab GTPase-activating protein Gyp1p localizes to the Golgi apparatus and is a negative regulator of Ypt1p. *Mol. Biol. Cell* **12**, 1215–1226 (2001).
- Strom, M., Vollmer, P., Tan, T. J. & Gallwitz, D. A yeast GTPase-activating protein that interacts specifically with a member of the Ypt/Rab family. *Nature* **361**, 736–739 (1993).
- Du, L. L., Collins, R. N. & Novick, P. J. Identification of a Sec4p GTPase-activating protein (GAP) as a novel member of a Rab GAP family. *J. Biol. Chem.* **273**, 3253–3256 (1998).
- Albert, S. & Gallwitz, D. Two new members of a family of Ypt/Rab GTPase activating proteins. Promiscuity of substrate recognition. *J. Biol. Chem.* **274**, 33186–33189 (1999).
- Cuif, M. H. *et al.* Characterization of GAPCenA, a GTPase activating protein for Rab6, part of which associates with the centrosome. *EMBO J.* **18**, 1772–1782 (1999).
- Vollmer, P., Will, E., Scheglmann, D., Strom, M. & Gallwitz, D. Primary structure and biochemical characterization of yeast GTPase-activating proteins with substrate preference for the transport GTPase Ypt7p. *Eur. J. Biochem.* **260**, 284–290 (1999).
- Lanzetti, L. *et al.* The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5. *Nature* **408**, 374–377 (2000).
- Will, E. & Gallwitz, D. Biochemical characterization of Gyp6p, a Ypt/Rab-specific GTPase-activating protein from yeast. *J. Biol. Chem.* **276**, 12135–12139 (2001).
- Gao, X. D. *et al.* The GAP activity of Msb3p and Msb4p for the Rab GTPase Sec4p is required for efficient exocytosis and actin organization. *J. Cell Biol.* **162**, 635–646 (2003).
- Haas, A. K., Fuchs, E., Kopajtich, R. & Barr, F. A. A. GTPase-activating protein controls Rab5 function in endocytic trafficking. *Nature Cell Biol.* **7**, 887–893 (2005).
- Miinea, C. P. *et al.* AS160, the Akt substrate regulating GLUT4 translocation, has a functional Rab GTPase-activating protein domain. *Biochem. J.* **391**, 87–93 (2005).
- Lafourcade, C., Galan, J. M., Gloor, Y., Haguener-Tsapis, R. & Peter, M. The GTPase-activating enzyme Gyp1p is required for recycling of internalized membrane material by inactivation of the Rab/Ypt GTPase Ypt1p. *Mol. Cell. Biol.* **24**, 3815–3826 (2004).
- Segev, N., Mulholland, J. & Botstein, D. The yeast GTP-binding YPT1 protein and a mammalian counterpart are associated with the secretion machinery. *Cell* **52**, 915–924 (1988).
- Saraste, J., Lahtinen, U. & Goud, B. Localization of the small GTP-binding protein rab1p to early compartments of the secretory pathway. *J. Cell Sci.* **108**, 1541–1552 (1995).
- Zheng, J. Y. *et al.* A novel Rab GTPase, Rab33B, is ubiquitously expressed and localized to the medial Golgi cisternae. *J. Cell Sci.* **111**, 1061–1069 (1998).
- Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. GTPase mechanism of G proteins from the 1.7-Å crystal structure of transducin α -GDP- AlF_4^- . *Nature* **372**, 276–279 (1994).
- Coleman, D. E. *et al.* Structures of active conformations of $G_{i\alpha 1}$ and the mechanism of GTP hydrolysis. *Science* **265**, 1405–1412 (1994).
- Mittal, R., Ahmadian, M. R., Goody, R. S. & Wittinghofer, A. Formation of a transition-state analog of the Ras GTPase reaction by Ras-GDP, tetrafluoroaluminate, and GTPase-activating proteins. *Science* **273**, 115–117 (1996).
- Scheffzek, K. *et al.* The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338 (1997).
- Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gamblin, S. J. Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature* **389**, 758–762 (1997).
- Nassar, N., Hoffman, G. R., Manor, D., Clardy, J. C. & Cerione, R. A. Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. *Nature Struct. Biol.* **5**, 1047–1052 (1998).
- Tesmer, J. J., Berman, D. M., Gilman, A. G. & Sprang, S. R. Structure of RGS4 bound to AlF_4^- -activated $G_{i\alpha 1}$: stabilization of the transition state for GTP hydrolysis. *Cell* **89**, 251–261 (1997).
- Slep, K. C. *et al.* Structural determinants for regulation of phosphodiesterase by a G protein at 2.0 Å. *Nature* **409**, 1071–1077 (2001).
- Daumke, O., Weyand, M., Chakrabarti, P. P., Vetter, I. R. & Wittinghofer, A. The GTPase-activating protein Rap1GAP uses a catalytic asparagine. *Nature* **429**, 197–201 (2004).
- Eathiraj, S., Pan, X., Ritacco, C. & Lambright, D. G. Structural basis of family-wide Rab GTPase recognition by rabenosyn-5. *Nature* **436**, 415–419 (2005).
- De Antoni, A., Schmitzova, J., Trepte, H. H., Gallwitz, D. & Albert, S. Significance of GTP hydrolysis in Ypt1p-regulated endoplasmic reticulum to Golgi transport revealed by the analysis of two novel Ypt1-GAPs. *J. Biol. Chem.* **277**, 41023–41031 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Novick for the *GYP1*-null strains and plasmids, and J. Saporita for assistance with yeast experiments. This work was supported by a National Institutes of Health grant.

Author Information Coordinates and structure factors for the Gyp1p-Rab33- AlF_3 complex have been deposited with the Protein Data Bank under the ID code 2G77. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.G.L. (david.lambright@umassmed.edu).

The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3

Paul A. C. Cloos^{1*}, Jesper Christensen^{1*}, Karl Agger^{1*}, Alessio Maiolica², Juri Rappsilber², Torben Antal¹, Klaus H. Hansen¹ & Kristian Helin^{1,3}

Methylation of lysine and arginine residues on histone tails affects chromatin structure and gene transcription^{1–3}. Tri- and dimethylation of lysine 9 on histone H3 (H3K9me3/me2) is required for the binding of the repressive protein HP1 and is associated with heterochromatin formation and transcriptional repression in a variety of species^{4–6}. H3K9me3 has long been regarded as a ‘permanent’ epigenetic mark^{7,8}. In a search for proteins and complexes interacting with H3K9me3, we identified the protein GASC1 (gene amplified in squamous cell carcinoma 1)⁹, which belongs to the JMJD2 (jumonji domain containing 2) subfamily of the jumonji family, and is also known as JMJD2C¹⁰. Here we show that three members of this subfamily of proteins demethylate H3K9me3/me2 *in vitro* through a hydroxylation reaction requiring iron and α -ketoglutarate as cofactors. Furthermore, we demonstrate that ectopic expression of GASC1 or other JMJD2 members markedly decreases H3K9me3/me2 levels, increases H3K9me1 levels, delocalizes HP1 and reduces heterochromatin *in vivo*. Previously, GASC1 was found to be amplified in several cell lines derived from oesophageal squamous carcinomas^{9,11,12}, and in agreement with a contribution of GASC1 to tumour development, inhibition of GASC1 expression decreases cell proliferation. Thus, in addition to identifying GASC1 as a histone trimethyl demethylase, we suggest a model for how this enzyme might be involved in cancer development, and propose it as a target for anti-cancer therapy.

Methylation of lysine 9 on histone H3 is associated with epigenetically silenced chromatin^{1–3}. As documented by studies of SUV39H1 and SUV39H2 knockout mice¹³, loss of the H3K9me3 mark results in genomic instability and predisposes to cancer. Hence, enzymes capable of reversing this particular mark have long been sought, although their existence has been questioned. The latter view has been reinforced by the fact that H3K9me3 is required for the establishment and maintenance of heterochromatin, a “very stable and heritable chromatin state”⁸. In the present study we show that the jumonji protein GASC1 (ref. 9) and its homologues JMJD2A and JMJD2B demethylate H3K9me3/me2.

In an effort to identify proteins that interact with H3K9me3, we performed affinity purification of HeLa nuclear proteins using H3 peptides either non-methylated or trimethylated at lysine 9. Several proteins were specifically enriched on H3K9me3 as compared to non-methylated H3K9 (Supplementary Fig. S1). Among the proteins identified was the jumonji protein GASC1 (Supplementary Fig. S2). Owing to the presence of a jumonji N domain (JmjN) and a jumonji C domain (JmjC), the protein is also denoted as JMJD2C¹⁰ (Supplementary Fig. S2). Of particular interest, JmjC domain proteins have been suggested to be histone demethylases¹⁴, and just before the

completion of this study the JmjC domain protein FBXL11 was shown to demethylate H3K36me2/me1 (ref. 15).

Because we found GASC1 to be associated with H3K9me3, we were intrigued by the idea that GASC1 might be involved in demethylating this epigenetic mark. To test this possibility we purified full-length, His-tagged human GASC1 from baculovirus-infected insect cells (Supplementary Fig. S3). Recombinant GASC1 was further purified by size-exclusion chromatography, and the highly purified GASC1 was tested for demethylase activity, by incubation with histones. As shown in Fig. 1a, GASC1 very efficiently reduced levels of H3K9me3. Similarly, GASC1 demethylated H3K9me3 on nucleosomes, the physiologically relevant template (Supplementary Fig. S4a).

To analyse further the specificity of the demethylase activity of GASC1, we incubated fractions from size-exclusion chromatography of GASC1 with histones and evaluated the methylation status of various epigenetic marks by immunoblotting (Fig. 1b and Supplementary Fig. S4b). Here, H3K9me3 and H3K9me2 from histones were completely removed in the fractions with the highest concentrations of GASC1 (Fig. 1b, fractions 9–12). A concomitant increase in the level of H3K9me1 was also observed, consistent with GASC1 converting H3K9me3 to H3K9me2 and then to H3K9me1. In contrast, the levels of other tested epigenetic marks including H3K4me3, H3K4me2, H3K27me3 and H4K20me3 were unaffected by GASC1 treatment (Fig. 1b and Supplementary Fig. S4b).

Notably, H3K9me2 appeared to be increased when histones were incubated with low to moderate amounts of GASC1 (Fig. 1b, fractions 6, 8, 14 and 16). The reason for the apparent increase in the H3K9me2 mark could be due to a steady-state level of this mark being reached in the initial phase of the demethylation reaction. Thus, it can be proposed that although some H3K9me2 is lost through GASC1 demethylation, additional H3K9me2 will be added to the global pool by GASC1 demethylation of H3K9me3, at least as long as significant levels of H3K9me3 are present. To confirm this assumption, we performed additional studies in which synthetic H3 peptide substrates methylated at K9 or K27 were incubated with varying amounts of GASC1. These experiments showed that even low concentrations of GASC1 could effectively remove both H3K9me3 and H3K9me2, whereas other tested methylated lysines (including H3K9me1) were unaffected (Supplementary Fig. S5).

Consistent with the results in Fig. 1a, b, the demethylase activity of GASC1 varies in a concentration-dependent manner and can be abolished by heat denaturation of GASC1 (Supplementary Fig. S4b), suggesting that it is an enzyme-dependent reaction. We also performed an additional *in vitro* demethylation study, where pure synthetic H3K9me3 peptide was incubated with GASC1. After

¹Biotech Research & Innovation Centre, Fruebjergvej 3, 2100 Copenhagen, Denmark. ²FIRC Institute of Molecular Oncology, Via Adamello 16, 20139 Milan, Italy. ³Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen, Denmark.

*These authors contributed equally to this work.

30 min incubation almost all H3K9me3 was converted to H3K9me2 and H3K9me1 (Supplementary Fig. S4c).

To demonstrate formally that GASC1 demethylates H3K9me3 and H3K9me2, mass spectrometry was performed on a synthetic H3K9me3 peptide incubated alone or in the presence of GASC1. This analysis revealed that most H3K9me3 peptide (approximately 80%) was converted to H3K9me2 and further to H3K9me1 after incubation with GASC1 (Fig. 1c, d).

Next, we considered whether the close GASC1 homologues JMJD2A and JMJD2B (Supplementary Fig. S2) could also demethylate H3K9me3/me2 *in vitro*. We cloned full-length human JMJD2A and JMJD2B and produced recombinant proteins from baculovirus-infected cells (Supplementary Fig. S3). Indeed, both homologues can demethylate H3K9me3/me2 (Supplementary Fig. S6).

To obtain an insight into the mechanism of the demethylation reaction, we modelled human GASC1 onto the structure of factor inhibiting HIF1 (FIH), the only jumonji protein for which the three-dimensional structure has been resolved^{16,17}. In agreement with the alignment (Supplementary Figs S7 and S8), the modelling indicated that residues H190, E192 and H288 form an essential part of the iron-binding groove of GASC1 (Fig. 1f). On the basis of this *in silico* model

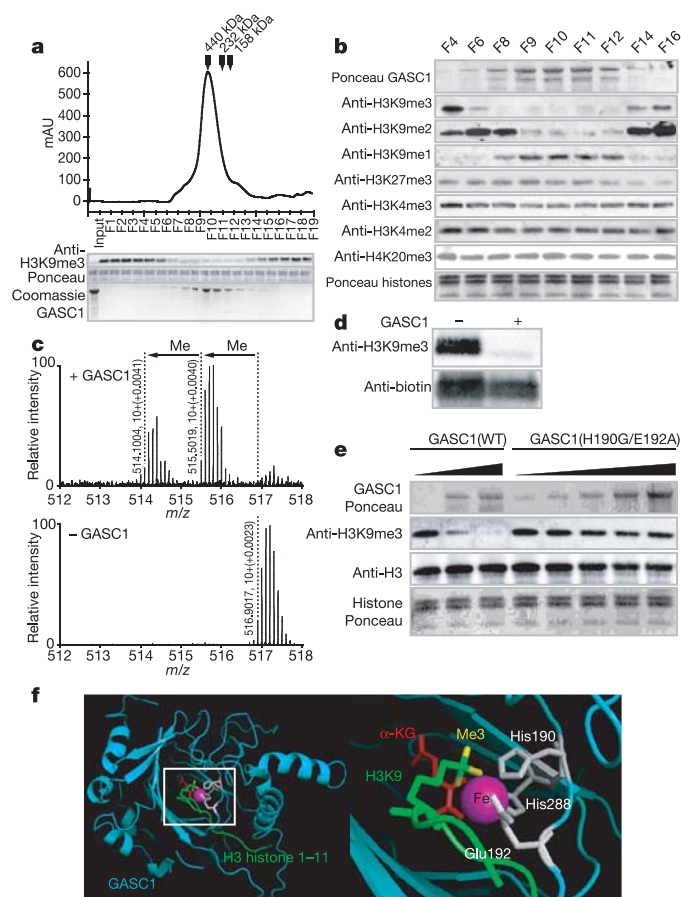


Figure 1 | GASC1 demethylates H3K9me3/me2. **a**, Fractions from size-exclusion chromatography of recombinant GASC1 were incubated with histones for 30 min at 37 °C and demethylation activity was assayed by immunoblotting. **b**, Demethylation activity in selected fractions assayed by immunoblotting. **c**, **d**, H3K9me3 peptides were incubated with or without GASC1 and the material was analysed by mass spectrometry (**c**) and immunoblotting (**d**). **e**, Demethylation assay using varying amounts of wild-type and mutant GASC1 on bulk histones. **f**, Predicted structure of GASC1 in complex with H3K9me3 tail (1–11) and its cofactors; the lysine H3K9 methyl groups are marked in yellow. Left panel shows an overview; right panel shows a magnification of the iron-binding residues. α -KG, α -ketoglutarate.

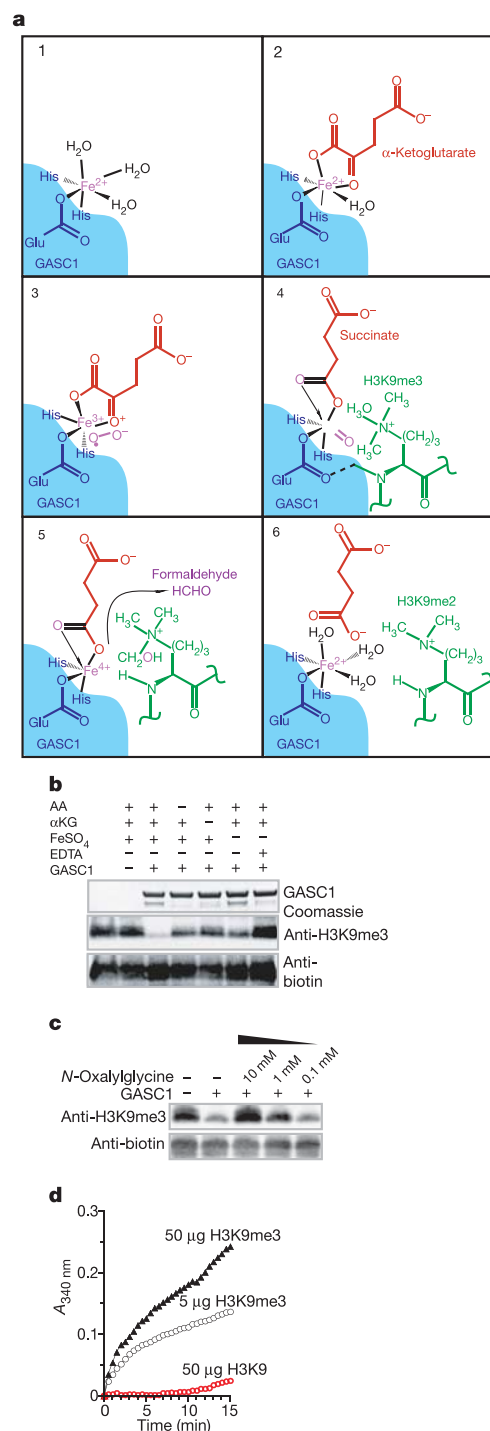


Figure 2 | Reaction mechanism for GASC1 demethylation. **a**, 1: Fe(II) is bound by residues His 190, Glu 192 and His 288. 2: α -Ketoglutarate is bound. 3: Iron binds oxygen. 4: Oxidative decarboxylation of α -ketoglutarate produces carbon dioxide, succinate and ferryl (Fe(IV) = O). 5: Ferryl hydroxylation of methylated K9 releases formaldehyde. 6: The histone tail and succinate leave the GASC1 molecule. **b**, Demethylation assay using a synthetic H3K9me3 peptide as substrate, in the presence or absence of cofactors. AA, ascorbic acid. KG, ketoglutarate. **c**, Demethylation assay using histones as substrates in the presence of varying amounts of *N*-oxalylglycine. **d**, Formaldehyde produced by incubating a fixed amount of GASC1 (40 μ g) with varying amounts of H3K9me3 peptide or the non-methylated peptide, measured using a demethylation–FDH-coupled assay²⁸. The y axis shows absorbance values at 340 nm.

we predicted that mutating H190 and E192 would be sufficient to abrogate the iron-binding ability of GASC1 and thus also to inhibit its demethylation activity.

Although wild-type GASC1 has a robust demethylase activity, mutant GASC1 has no detectable demethylase activity (Fig. 1e), thereby indicating that the putative iron-binding residues are crucial to the demethylase activity of GASC1.

Dioxygenases belonging to the cupin superfamily are dependent upon Fe(II) and α -ketoglutarate^{16,18}. In addition, some dioxygenases have the additional requirement of ascorbate for full catalytic activity^{15,19,20}.

To test whether GASC1-mediated demethylation would fit the reaction mechanism described above and depicted in Fig. 2a, we first tested the importance of the putative cofactors for the demethylation reaction.

In the presence of all its cofactors GASC1 demethylates H3K9me3 *in vitro*, whereas GASC1 incubated with EDTA (chelating iron) or in the absence of its cofactors was significantly less efficient (Fig. 2b).

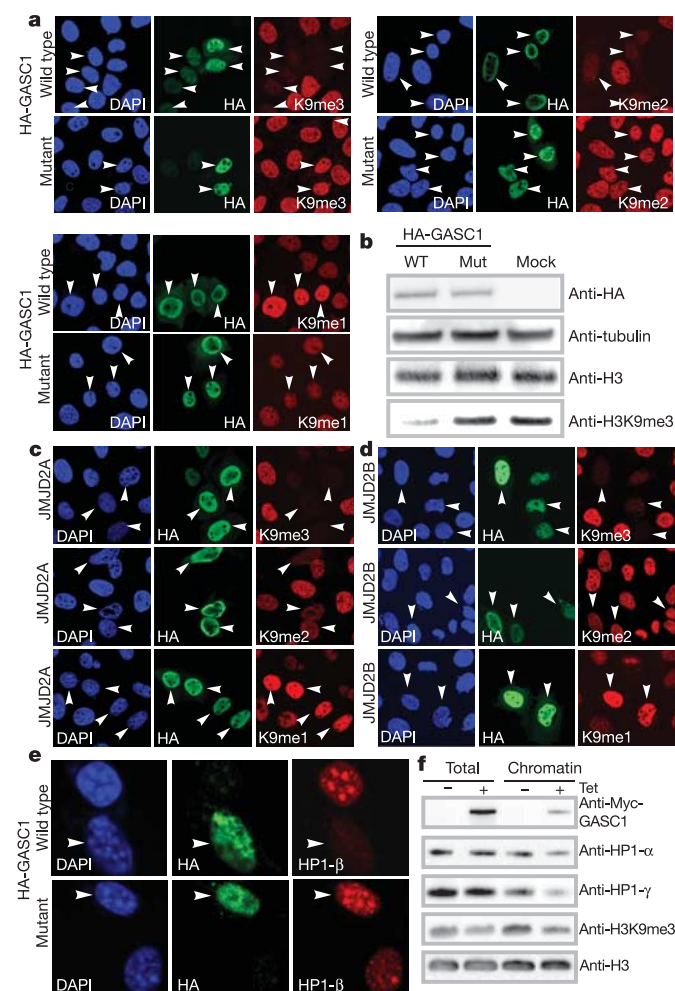


Figure 3 | Ectopic expression of GASC1 leads to loss of H3K9me3/me2 *in vivo*. **a**, Demethylation of H3K9me3/me2 by wild-type but not mutant GASC1 *in vivo* as measured by immunofluorescence. White arrowheads indicate transfected cells. **b**, Effect of ectopic expression of haemagglutinin (HA)-tagged GASC1 and mutant on H3K9me3 *in vivo* on histones. **c**, **d**, Ectopic expression of JMJD2A and JMJD2B leads to demethylation of H3K9me3 and H3K9me2 and an increase in H3K9me1 levels in U2OS cells. **e**, Confocal microscopy of mouse embryo fibroblasts transfected with HA-tagged GASC1. HP1- β is delocalized in cells overexpressing GASC1. **f**, Loss of chromatin-associated HP1- α and HP1- γ in HEK293 cells expressing tetracycline (Tet)-inducible Myc-tagged GASC1. Total indicates total lysate, whereas chromatin indicates the chromatin fraction.

The ability of the demethylation reaction to occur *in vitro* without the addition of cofactors is probably due to the co-purification of cofactors (for example, iron and α -ketoglutarate) with recombinant GASC1. For this reason we sought to establish further the importance of these cofactors by testing the ability of *N*-oxalylglycine, CoCl₂ and NiSO₄ to inhibit GASC1-mediated demethylation of H3K9me3. These compounds are all able to inhibit the demethylation reaction effectively (Fig. 2c and Supplementary Fig. S9). *N*-oxalylglycine is an α -ketoglutarate analogue and presumably inhibits the activity of GASC1 by displacing α -ketoglutarate from the iron-binding residues of GASC1. Similarly, inhibition of GASC1 by CoCl₂ and NiSO₄ probably involves dislocation of Fe(II) from the iron-binding site of

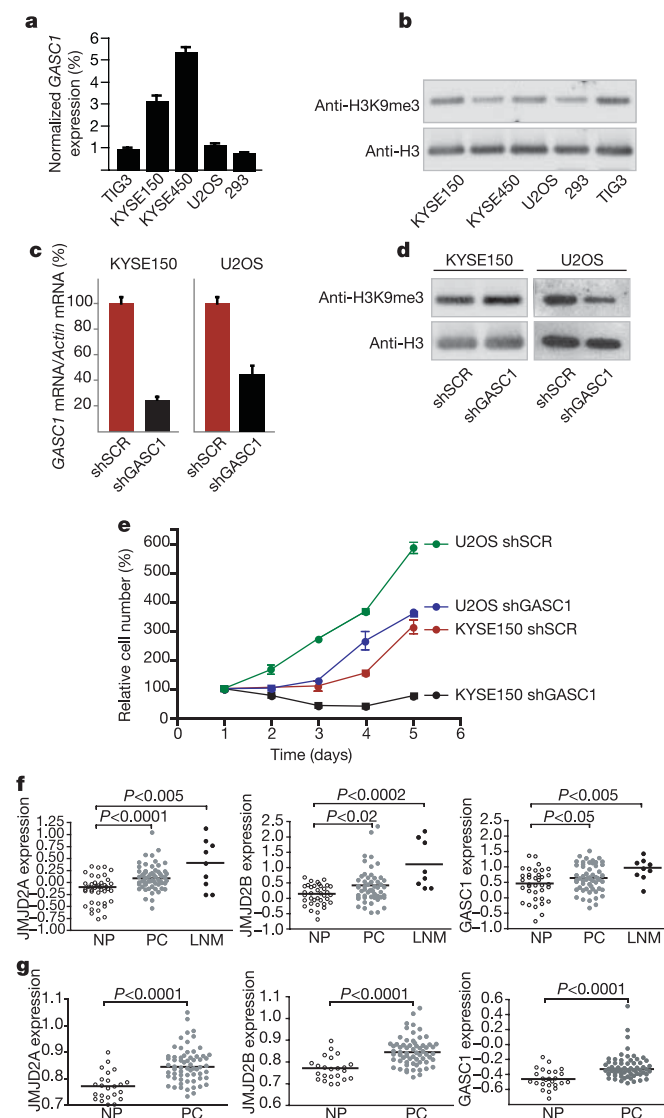


Figure 4 | Inhibition of GASC1 leads to decreased proliferation. **a**, GASC1 expression in various cell lines determined by real-time quantitative PCR (RT-qPCR). **b**, H3K9me3 status in various cell lines. **c**, KYSE150 and U2OS cells were transfected with shRNA targeting GASC1 or control shRNA. The efficiency of the shRNA interference was evaluated by RT-qPCR. **d**, **e**, H3K9me3 status and growth curves of KYSE150 and U2OS cells after shRNA expression. **f**, **g**, Data from ref. 29 (**f**) and ref. 30 (**g**) re-analysed to show expression levels of JMJD2 proteins in prostate tumours (PC), normal prostate tissues (NP) and unmatched lymph node metastases (LNM). Bars indicate medians; *P*-values of Mann-Whitney *U*-test are provided; see Supplementary Information for further details. Error bars on qPCR experiments and growth curves indicate standard deviations and standard error of the mean, respectively.

GASC1 by cobalt(II) or nickel(II) ions. The requirement of ascorbic acid, α -ketoglutarate and iron in the demethylation reaction and the inhibition of the reaction by nickel and cobalt salts and *N*-oxalylglycine strongly indicate that GASC1 is a dioxygenase.

Moreover, in agreement with the proposed reaction scheme, incubation of GASC1 with its substrate released formaldehyde (Fig. 2d). Taken together, these results demonstrate that GASC1 directly demethylates H3K9me3 and H3K9me2 *in vitro* through a hydroxylation reaction requiring iron and α -ketoglutarate, producing formaldehyde and H3K9me1.

Having established that GASC1 can demethylate H3K9me3 and H3K9me2 *in vitro*, we next considered whether GASC1 could modulate heterochromatin formation/maintenance *in vivo*. Ectopic expression of GASC1 in human and murine fibroblasts caused an efficient decrease of H3K9me3 (Fig. S10a,b) and an increase of H3K9me1 (Supplementary Fig. S10a).

The decrease in H3K9me3/me2 levels and increase of H3K9me1 was also apparent in U2OS cells transfected with wild-type GASC1, but not with the putative Fe(II)-binding mutant (Fig. 3a,b), further confirming the crucial importance of these predicted Fe(II)-coordinating residues. In contrast, levels of H3K4me3 and H3K27me3 were not affected (Supplementary Fig. S11).

As observed for GASC1, ectopic expression of JMJD2A and JMJD2B led to a significant reduction of H3K9me3 and H3K9me2 levels, with a concomitant increase in H3K9me1 (Fig. 3c, d). These results suggest that GASC1 belongs to a family of histone H3K9 demethylases.

Heterochromatin formation and maintenance requires the presence of H3K9me3/me2 and HP1 binding^{5,6,8}. Because the ectopic expression of GASC1 leads to a global reduction of H3K9me3/me2 levels, we tested whether increased amounts of GASC1 affected HP1 binding and localization. As predicted, HP1- β was delocalized in NIH 3T3 cells overexpressing GASC1 (Fig. 3e). Similarly, ectopic expression of the GASC1 homologues JMJD2A and JMJD2B led to delocalization of HP1- β (Supplementary Fig. S12). We further validated these results by demonstrating decreased levels of HP1- α and HP1- γ associated with chromatin in GASC1-overexpressing cells (Fig. 3f). Taken together, these results suggest that GASC1 might have a physiologically important role in controlling heterochromatin formation and maintenance.

The strict control of heterochromatin formation and maintenance is critical for both proper biological function and genomic integrity. For instance, centromeric heterochromatin formation is essential for the correct segregation of chromosomes during mitosis²¹. Similarly, it has been demonstrated that the deletion of *Suv39h1* and *Suv39h2* genes in mice, the products of which are responsible for di- and trimethylation of H3K9 in heterochromatin, leads to chromosomal instability and collaborates with oncogenes in inducing mouse lymphomas¹³. Therefore, aberrant activity of GASC1 or its homologues could potentially lead to genomic instability and consequently cancer. Indeed, GASC1 was originally identified as a gene frequently amplified in oesophageal squamous cell carcinoma^{9,12,22} and GASC1 is overexpressed in various cancer types containing chromosomal aberrations¹¹.

To begin to elucidate the physiological function of GASC1 in normal and cancer cells, we tested the relative levels of GASC1 mRNA and the H3K9me3 mark in two oesophageal carcinoma cell lines that have been reported to contain several copies of the GASC1 gene (KYSE150 and KYSE450)⁹. The human cell lines U2OS, 293 and TIG3 were included for comparison. As shown in Fig. 4a, GASC1 levels are indeed increased 3–5-fold in the oesophageal carcinoma cell lines as compared to other carcinoma cell lines and normal human fibroblasts. However, the increased expression of GASC1 did not lead to a global decrease in H3K9me3 levels (Fig. 4b). These results demonstrate that the increased levels of GASC1 are not sufficient to lead to global demethylation of H3K9me3, and may imply that GASC1 expressed at physiological levels regulates the H3K9me3 levels locally.

To test whether GASC1 can contribute to tumour cell maintenance, we inhibited the expression of GASC1 in KYSE150 and (as a control) U2OS cells by short hairpin RNA (shRNA). The efficiency of inhibition of GASC1 expression was similar in the two cell lines (Fig. 4c), and consistent with the results reported above, we did not observe changes in the global levels of H3K9me3 (Fig. 4d). Notably, however, inhibition of GASC1 expression caused a significant reduction of proliferation in both KYSE150 and U2OS cells (Fig. 4e). Although the inhibition of cell proliferation was more pronounced in KYSE150 cells, this finding indicates that GASC1 is required for cell proliferation, also of cells in which GASC1 levels are normal. Therefore, future studies are required to understand whether GASC1 contributes to tumour development.

Meanwhile, to obtain further support for a potential involvement of GASC1 in cancer, we searched the Oncomine database²³ for differential GASC1 expression in normal versus tumour tissue. The expression of GASC1 and its homologues JMJD2A and JMJD2B were significantly increased in prostate cancers relative to normal tissue (Fig. 4f, g). These results suggest that increased expression of all three JMJD2 proteins might contribute to the development of human cancer possibly by demethylation of H3K9me3/me2.

Di- and trimethylated H3K9 is required for HP1 binding, essential for heterochromatin formation and is associated with transcriptional repression. Moreover, these epigenetic marks are increased on specific genes in senescence, a key mechanism guarding cells against cancer^{24–27}. It can therefore be speculated that amplification of GASC1 may reduce the ability of cells to become senescent, or lead to de-repression of otherwise silenced oncogenes. Accordingly, inhibition of GASC1-demethylating activity could potentially constitute a new anti-neoplastic therapeutic modality.

Using several independent lines of evidence we have shown that GASC1 and its homologues JMJD2A and JMJD2B demethylate the repressive histone H3K9me3/me2 marks both *in vitro* and *in vivo*. Our findings demonstrate that histone trimethylation is a reversible modification. This finding may potentially have far-reaching implications for human disease, notably cancer. Further studies will be required to unravel the biology of GASC1 and the jumonji protein family and shed light on their possible role in epigenetic signalling, transcriptional regulation and human disease.

METHODS

Demethylation assay. Bulk histones or synthetic histone peptides were incubated with purified His-tagged GASC1 in demethylation buffer (50 mM Tris pH 8.0, 50 mM KCl, 10 mM MgCl₂, 1 mM α -ketoglutarate, 40 μ M FeSO₄, 2 mM ascorbic acid) at 37 °C. When using nucleosomes as substrate MgCl₂ was omitted in the reaction. In a typical reaction, either 6 μ g of bulk histones or 2 μ g of modified histone peptides were incubated with 20 μ g GASC1 in a volume of 100 μ l for 30 min. Reaction mixtures were analysed by either western blotting using specific antibodies, or by formaldehyde release assays.

Formaldehyde release assay. Formaldehyde release assays were performed essentially as described²⁸. All reactions were performed in a total volume of 200 μ l per reaction in a quartz cuvette. In short, recombinant GASC1 (typically 40 μ g) was incubated in 150 μ l demethylation buffer (see above) in the presence of 2 mM NAD⁺ and 0.2 U formaldehyde dehydrogenase (FDH) at 37 °C for 5 min. Then the reaction was started by adding the substrates (histone peptides). The absorbance at 340 nm was measured with 0.5-min intervals (15 min total) using a Genesys 10UV Thermospectronic spectrophotometer at 37 °C.

GASC1 demethylation of H3K9me3 peptide. Six micrograms of recombinant GASC1 was incubated with 3 μ g H3K9me3 peptide in FDH buffer in a final volume of 90 μ l for 30 min at 37 °C. Urea was added to a final concentration of 4 M and the mixture was incubated at 20 °C for 15 min. An equal volume of 1% trifluoroacetic acid (TFA) was added and the sample loaded on a reversed-phase mini C8 column packed in a 100- μ l tip (column volume of 20 μ l). After washing in 1% TFA, the bound peptide was eluted in 20 μ l (30% methanol, 25% formic acid). One-third of the eluted material was analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and western blotting using first anti-H3K9me3, followed by anti-biotin antibody. The rest of the material was analysed by mass spectrometry.

Mass spectrometry analysis. One-third of the eluate was injected in 1% TFA using an Agilent 1100 Nano HPLC onto a C18 column (Reprosil-Pur C18-AQ

3 μm ; Dr. Maisch GmbH) packed into a spray emitter (75 μm internal diameter, 8 μm opening, 70 mm length; New Objectives). Peptides were eluted in a gradient from buffer A (5% acetonitrile and 0.5% acetic acid) to buffer B (acetonitrile and 0.5% acetic acid) going from 0% to 20% in 10 min at 300 nl min⁻¹. Spectra were recorded on a LTQ-FT mass spectrometer (ThermoFinnigan).

Additional materials and methods, including methods for cloning procedures and immunostaining, are described in the Supplementary Methods.

Received 13 January; accepted 27 April 2006.

Published online 28 May 2006.

- Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**, 172–183 (2003).
- Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).
- Peterson, C. L. & Laniel, M. A. Histones and histone modifications. *Curr. Biol.* **14**, R546–R551 (2004).
- Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
- Elgin, S. C. & Grewal, S. I. Heterochromatin: silence is golden. *Curr. Biol.* **13**, R895–R898 (2003).
- Kubicek, S. & Jenuwein, T. A crack in histone lysine methylation. *Cell* **119**, 903–906 (2004).
- Bannister, A. J. & Kouzarides, T. Reversing histone methylation. *Nature* **436**, 1103–1106 (2005).
- Yang, Z. Q. *et al.* Identification of a novel gene, GASC1, within an amplicon at 9p23–24 frequently detected in esophageal cancer cell lines. *Cancer Res.* **60**, 4735–4739 (2000).
- Katoh, M. & Katoh, M. Identification and characterization of JMJD2 family genes *in silico*. *Int. J. Oncol.* **24**, 1623–1628 (2004).
- Yang, Z. Q. *et al.* A novel amplicon at 9p23–24 in squamous cell carcinoma of the esophagus that lies proximal to GASC1 and harbors NF1B. *Jpn. J. Cancer Res.* **92**, 423–428 (2001).
- Hu, N. *et al.* Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. *Cancer Res.* **65**, 2542–2546 (2005).
- Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
- Trewick, S. C., McLaughlin, P. J. & Allshire, R. C. Methylation: lost in hydroxylation? *EMBO Rep.* **6**, 315–320 (2005).
- Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).
- Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 α . *J. Biol. Chem.* **278**, 1802–1806 (2003).
- Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M. & Ryu, S. E. Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *J. Biol. Chem.* **278**, 7558–7563 (2003).
- Hegg, E. L. & Que, L. Jr. The 2-His-1-carboxylate facial triad—an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **250**, 625–629 (1997).
- Knowles, H. J., Raval, R. R., Harris, A. L. & Ratcliffe, P. J. Effect of ascorbate on the activity of hypoxia-inducible factor in cancer cells. *Cancer Res.* **63**, 1764–1768 (2003).
- Kivirikko, K. I. & Myllyharju, J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol.* **16**, 357–368 (1998).
- Pidoux, A. L. & Allshire, R. C. The role of heterochromatin in centromere function. *Phil. Trans. R. Soc. Lond. B* **360**, 569–579 (2005).
- Ragvin, A. *et al.* Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.* **337**, 773–788 (2004).
- Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
- Braig, M. *et al.* Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature* **436**, 660–665 (2005).
- Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).
- Collado, M. *et al.* Tumour biology: senescence in premalignant tumours. *Nature* **436**, 642 (2005).
- Michaloglou, C. *et al.* BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
- Shi, Y. J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19**, 857–864 (2005).
- Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA* **101**, 811–816 (2004).
- Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22**, 2790–2799 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements We are grateful to U. Toftegaard and S. Keshtar for technical assistance. We thank A. Bracken, A. H. Lund and C. Storgaard Sørensen for critical reading of the manuscript. We thank members of the Helin laboratory and M. Salek for technical advice and support. This work was supported by grants from the Danish Cancer Society, the Novo Nordisk Foundation, the Danish Medical Research Council, the Danish Natural Science Research Council and the Danish Ministry of Science, Technology and Innovation.

Author Contributions P.A.C.C. performed the *in vitro* binding experiments identifying GASC1 as an H3K9me3 interactor, suggested that GASC1 could be an H3K9me3-specific demethylase and co-wrote the paper. J.C. performed most DNA cloning steps and produced recombinant proteins. P.A.C.C. and J.C. established the *in vitro* demethylation assays, identified GASC1 as an H3K9me3-specific demethylase and performed various biochemical experiments. K.A. performed the *in vivo* experiments and immunofluorescence studies. K.H.H. designed the peptides and implemented the *in vitro* binding assays, which identified H3K9me3 interactors, and performed various biochemical experiments including preparation of samples for mass spectrometry. A.M. and J.R. performed mass spectrometry. T.A. performed *in silico* modelling studies. K.H. suggested the strategy to identify proteins binding to the modified histone tails, contributed to the planning of experiments and co-wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.H. (kristian.helin@bric.dk).

LETTERS

The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36

Robert J. Klose^{1,2}, Kenichi Yamane^{1,2}, Yangjin Bae³, Dianzheng Zhang³, Hediye Erdjument-Bromage⁴, Paul Tempst⁴, Jiemin Wong³ & Yi Zhang^{1,2}

Post-translational modification of chromatin has profound effects on many biological processes including transcriptional regulation, heterochromatin organization, and X-chromosome inactivation^{1,2}. Recent studies indicate that methylation on specific histone lysine (K) residues participates in many of these processes³. Lysine methylation occurs in three distinct states, having either one (me1), two (me2) or three (me3) methyl groups attached to the amine group of the lysine side chain. These differences in modification state have an important role in defining how methylated chromatin is recognized and interpreted^{4–6}. Until recently, histone lysine methylation was considered a stable modification^{7,8}, but the identification of histone demethylase enzymes has demonstrated the reversibility of this epigenetic mark^{9–11}. So far, all characterized histone demethylases show enzymatic activity towards lysine residues modified in the me1 or me2 state^{9–11}, leaving open the possibility that me3 constitutes an irreversible modification. Here we demonstrate that JHDM3A (jumonji C (JmjC)-domain-containing histone demethylase 3A; also known as JMJD2A) is capable of removing the me3 group from modified H3 lysine 9 (H3K9) and H3 lysine 36 (H3K36). Overexpression of JHDM3A abrogates recruitment of HP1 (heterochromatin protein 1) to heterochromatin, indicating a role for JHDM3A in antagonizing methylated H3K9 nucleated events. siRNA-mediated knockdown of JHDM3A leads to increased levels of H3K9 methylation and upregulation of a JHDM3A target gene, *ASCL2*, indicating that JHDM3A may function in euchromatin to remove histone methylation marks that are associated with active transcription¹².

An extensive family of genes that encode JmjC-domain-containing proteins have been identified in eukaryotic genomes from yeast to human^{13,14}. Using an unbiased biochemical approach, we have previously identified the JmjC domain as a signature motif for enzymes that catalyse histone demethylation^{10,11}. In an attempt to identify additional histone lysine demethylases, we compared the JmjC domains of other JmjC family members to the known histone demethylases, JHDM1A (also known as FBXL11), JHDM1B (FBXL10) and JHDM2A (JMJD1A)^{10,11}. Of the JmjC-domain-containing proteins analysed, the JMJD2 family of proteins¹⁵ were excellent candidates for demethylase activity because amino acids predicted to be associated with enzymatic activity through their involvement in binding of Fe(II) and α -ketoglutarate (α -KG) are conserved (Supplementary Fig. S1a). As JMJD2A is the only JMJD2 family member characterized^{16,17}, we have focused our efforts in defining whether this protein is an active histone demethylase.

To determine whether JMJD2A is an active histone demethylase,

we incubated recombinant Flag-tagged JMJD2A produced in insect cells (Supplementary Fig. S1b) with histone substrates containing radioactively labelled methyl groups corresponding to several characterized methyl-lysine sites on histones H3 and H4. The results presented in Fig. 1a demonstrate that demethylation reactions containing Dim5-labelled histone substrate showed substantial release of labelled formaldehyde. A small, yet reproducible, release of labelled formaldehyde was also observed in reactions containing SET2-labelled substrate (Fig. 1a). These data suggest that JMJD2A may potentially demethylate modified H3K9 and H3K36 residues. To further refine the substrate and methylation-state specificity of JMJD2A, a demethylase assay was carried out on purified native core histones and the modification state of individual methylation sites was analysed by western blot (Fig. 1b and Supplementary Fig. S2a). This analysis revealed that JMJD2A efficiently demethylates both H3K9me3 and H3K36me3 resulting in an accumulation of H3K9me1 and H3K36me1, whereas the levels of other histone methylation sites remained unchanged (Supplementary Fig. S2a). Furthermore, JMJD2A only functions efficiently to demethylate H3K9 and H3K36 within core histone, but not nucleosomal, substrates (Supplementary Fig. S2b), which explains why SET2-labelled nucleosomal histone is not a good substrate in the radioactive formaldehyde-release assay (Fig. 1a). These data verify that JMJD2A is an active histone demethylase and demonstrate that the trimethyl-lysine state is enzymatically reversible. Because JMJD2A is the third JmjC-domain-containing histone demethylase, we have named the protein JHDM3A (JmjC-domain-containing histone demethylase 3A) to reflect its enzymatic function and conform to our existing naming convention.

To confirm that JHDM3A uses an oxidative demethylation mechanism with Fe(II) and α -KG as cofactors, each cofactor was independently omitted from the demethylation reaction. Full enzymatic activity was observed when the complete cofactor/enzyme complement was present in the reaction (Fig. 1c). Omitting Fe(II) from the reaction buffer did not result in loss of activity, suggesting that Fe(II) remains associated with the enzyme during the purification process—consistent with the fact that addition of the divalent chelator EDTA abrogated the enzymatic activity (data not shown). Ascorbate was required for enzymatic activity, presumably owing to its capacity to reduce Fe(III) to Fe(II). Together, these data demonstrate that JHDM3A is an oxidative histone demethylase with the capacity to directly reverse H3K9me3 and H3K36me3 modifications.

To further define the modification-state specificity of JHDM3A, we used methylated peptide substrates in demethylation reactions

¹Howard Hughes Medical Institute, and ²Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA. ³Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

⁴Molecular Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

coupled with mass spectrometry (Fig. 1d–g). In agreement with observations made using core histone substrates, JHDM3A is capable of demethylating H3K9me3 and H3K36me3 peptides (Fig. 1d, f). Notably, the H3K9me3 peptide was more extensively demethylated (~80%) than the H3K9me2 (~30%), H3K36me3 (~35%) or H3K36me2 (~25%) peptides, indicating that this is the preferred modification state targeted by JHDM3A (Fig. 1d–g). Demethylation by JHDM3A is not highly processive, as the extent of demethylation becomes less pronounced with the removal of each consecutive methyl group and does not culminate in the production of a single modification state. Furthermore, mass spectrometric analysis demonstrates that JHDM3A has the capacity to actively demethylate

all three methylation states (Fig. 1d–g). In combination, our *in vitro* analysis demonstrates that JHDM3A can directly remove methyl groups from modified H3K9 and H3K36 to produce H3K9/H3K36 having me2, me1 and me0 modifications.

Next, we sought to determine whether JHDM3A removes H3K9 and H3K36 modifications *in vivo*. To this end, Flag-tagged JHDM3A was overexpressed in cultured cells and H3K9/H3K36 methylation levels were analysed by indirect immunofluorescence (Fig. 2a–f and Supplementary Figs S4–S6) and western blotting (Fig. 2g). In agreement with the *in vitro* substrate specificity demonstrated for JHDM3A, we observed a near-complete loss of H3K9me3/H3K36me3 methylation in cells expressing JHDM3A (Fig. 2a, d, g, and Supplementary Fig. S5–S6). This activity is dependent on an intact JmjC domain, as an amino-acid substitution (H188A) in the predicted Fe(II)-binding site abrogates these effects (Fig. 2a, d, g, and Supplementary Figs S5–S6). Overexpression of JHDM3A resulted in an increase in H3K9me1-staining intensity in approximately one-third of transfected cells, and caused a subtle increase in H3K9me1 levels as assessed by western blotting (Fig. 2c, g, and Supplementary Fig. S7), but no changes in the level of H3K9me2 were observed (Fig. 2b, g). In addition, H3K36me2 levels were reduced and H3K36me1 levels increased in JHDM3A-expressing cells (Fig. 2e–g). Therefore, overexpression of JHDM3A *in vivo* causes dramatic loss of H3K9me3/H3K36me3 leading to accumulation of H3K9me1/H3K36me1.

Next, we attempted to determine which domains are required for demethylase activity *in vivo*. To achieve this, we generated mutants of JHDM3A harbouring deletions of predicted functional domains (Fig. 2h) and analysed their effect on H3K9me3/H3K36me3 staining in mouse cells. Both the JmjN and JmjC domains are required for demethylase activity (Fig. 2i, j), and deletion of the JmjN domain or the PHD and Tudor domains resulted in both nuclear and cytoplasmic localization of JHDM3A, indicating that these domains contribute to normal nuclear accumulation. Given that the PHD and Tudor domains are dispensable for enzymatic activity, these domains are probably involved in protein localization or protein–protein interaction.

Because little is known about the function of H3K36me modification in mammals¹⁸, we sought to examine in more detail the effects that JHDM3A has on H3K9me-mediated cellular events. Heterochromatin protein 1 (HP1)- α and - β preferentially bind to methylated H3K9 *in vitro*^{5,19,20} and localize to pericentric heterochromatin in an H3K9me2/3-dependent manner^{21,22}. Therefore, we proposed that JHDM3A activity may influence HP1 localization patterns. Consistent with this prediction, cells expressing high levels of JHDM3A show redistribution of pericentric HP1- α /HP1- β (Fig. 3a, b (middle panels) and Supplementary Fig. S8), and this effect is dependent on functional JHDM3A demethylase activity (Figs 3a, b (right panels) and Supplementary Fig. S8). Notably, HP1 localization is retained in a proportion of cells expressing JHDM3A, indicating that factors in addition to H3K9me also contribute to HP1 accumulation at pericentric heterochromatin (Supplementary Fig. S8). These data indicate that elevated levels of JHDM3A can function to antagonize HP1 recruitment to pericentric heterochromatin.

To examine whether JHDM3A has a role in removing histone methylation at a euchromatic gene, we used short interfering RNA (siRNA)-mediated knockdown to manipulate JHDM3A levels and analysed its effects on the only known JHDM3A target gene, *ASCL2* (ref. 17). Consistent with JHDM3A functioning as a negative regulator of transcription and an H3K9/H3K36 demethylase, siRNA-mediated knockdown resulted in *ASCL2* upregulation (Fig. 4b, c), and an increase in the levels of H3K9me3 and H3K36me2 at the *ASCL2* locus as assessed by radioactive polymerase chain reaction (PCR; Fig. 4d). In contrast, no significant changes in methylation were observed at the human chromosome 4 pericentric region (Fig. 4e). To quantify the changes in histone modification, we

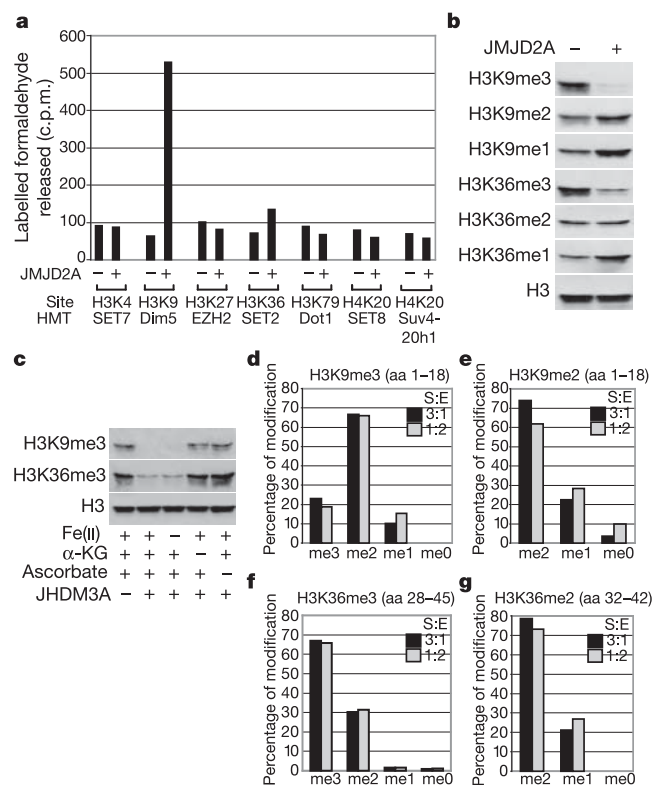


Figure 1 | JMJD2A/JHDM3A is a histone H3K9/H3K36 demethylase capable of removing trimethyl-lysine. **a**, Histones were labelled using various histone methyltransferase (HMT) enzymes as indicated below the bar graph, and equal counts of the various substrates were incubated with recombinant JMJD2A. The release of labelled methyl groups was measured to assay for histone demethylase activity, in the presence (+) or absence (–) of enzyme. JMJD2A demethylates H3K9 labelled by the histone methyltransferase Dim5, and to a lesser extent H3K36 labelled by SET2. **b**, HeLa cell core histones were incubated in the presence (+) or absence (–) of JMJD2A, and following the demethylation reaction histone methylation levels were analysed by western blotting with modification-specific antibodies. JMJD2A demethylates H3K9me3/H3K36me3 resulting in an accumulation of H3K9me1/H3K36me1. For the reasons stated in the text, JMJD2A will hereafter be referred to as JHDM3A. **c**, To demonstrate that JHDM3A carries out demethylation using the predicted oxidative mechanism, the proposed cofactors were omitted from the reaction as indicated (–) below the western blot. Full enzymatic activity relies on the presence of the cofactor α -KG and ascorbate in the reaction. **d–g**, H3K9me3, H3K9me2, H3K36me3 and H3K36me2 peptides were incubated with JHDM3A at substrate to enzyme (S:E) ratios of 3:1 (black bars) and 1:2 (grey bars) in demethylase assays before mass spectrometry. The bar graph represents the percentage of each modification state after the demethylation reaction. The original mass spectrometry data for the S:E ratio of 1:2 is presented in Supplementary Fig. S3. JHDM3A demethylates H3K9me3/2 and H3K36me3/2 peptides, showing the highest activity towards H3K9me3.

used real-time PCR following siRNA-mediated knockdown using two independent siRNA molecules targeting *JHDM3A* (Fig. 4f). This quantitative analysis revealed an approximate fourfold increase in H3K9me3 levels after siRNA treatment, but yielded more subtle changes in H3K36me2 levels than were observed by radioactive PCR, perhaps owing to lower levels of H3K36 methylation at the JAR (for 'JHDM3A-associated region') site. Therefore, occupancy of JHDM3A at a specific target gene is required to maintain reduced levels of H3K9 methylation and, to a lesser extent, H3K36 methylation, in concert with transcriptional repression.

During the revision of our manuscript, another study reported

that JHDM3A is a me3-specific H3K9/H3K36 demethylase, with little activity towards lower modification states²³. In contrast, we clearly observe that JHDM3A demethylates H3K9/H3K36me2/me1. In our hands, JHDM3A shows poor processivity under suboptimal reaction conditions *in vitro*, perhaps providing an explanation for the failure of this previous report to observe significant activity of JHDM3A towards me2/me1 modification states *in vitro*. We are confident that JHDM3A can demethylate lower H3K9/H3K36 methylation states, as several independent *in vitro* and *in vivo* lines of evidence support this conclusion (Fig. 1b, d–g, Supplementary Fig. S2a, Fig. 2a–h and Fig. 4d).

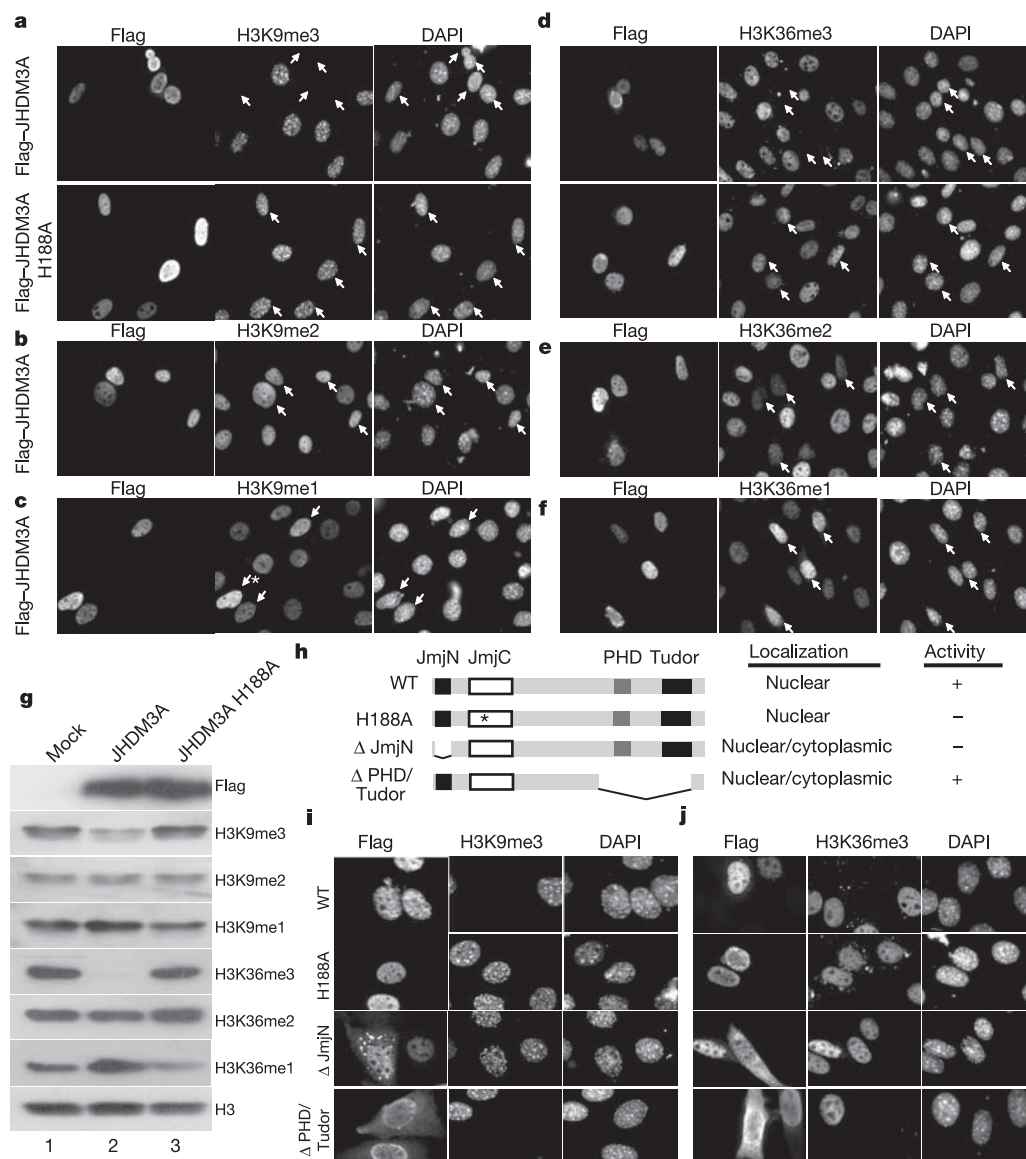


Figure 2 | JHDM3A requires the JmjC and JmjN domains to demethylate H3K9/H3K36 *in vivo*. **a–f**, JHDM3A was expressed in mouse NIH3T3 cells as a Flag fusion protein. Indirect immunofluorescence with antibodies against Flag (left panels) or methylated histone (middle panels) were used to analyse the substrate specificity of JHDM3A *in vivo*. DAPI (4,6-diamidino-2-phenylindole) staining (right panels) indicates location of nuclei in each field. Cells transfected with JHDM3A (arrows) showed a pronounced loss of H3K9me3/H3K36me3 staining (**a** and **d**, top panels), and this activity was dependent on an intact JmjC domain as methyl-histone staining was unaffected when a mutation in the predicted Fe(II)-binding site was introduced into JHDM3A (**a** and **d**, bottom panels). JHDM3A-mediated demethylation results in an accumulation of H3K9me1/H3K36me1 signal (the asterisk indicates a cell showing increased levels of H3K9me1) (**c** and **f**).

The H3K9me2/H3K36me2 signal is also shown (**b** and **e**). **g**, JHDM3A was expressed in HEK293T cells as a Flag fusion protein, and histones were analysed by western blotting with antibodies against H3K9me/H3K36me modifications. A pronounced loss of H3K9me3/H3K36me3 was observed in JHDM3A-expressing cells (lane 2), and this was dependent on an intact JmjC domain (lane 3). A subtle increase in H3K9me1 and a clear increase in H3K36me1 levels were also evident in JHDM3A-expressing cells.

h–j, JHDM3A wild-type (WT) and deletion proteins (**h**) were expressed in NIH3T3 cells and indirect immunofluorescence was used to analyse H3K9me3/H3K36me3 levels (**i** and **j**). The enzymatic activity of JHDM3A requires both the JmjC and JmjN domains (compare middle panels to top and bottom panels).

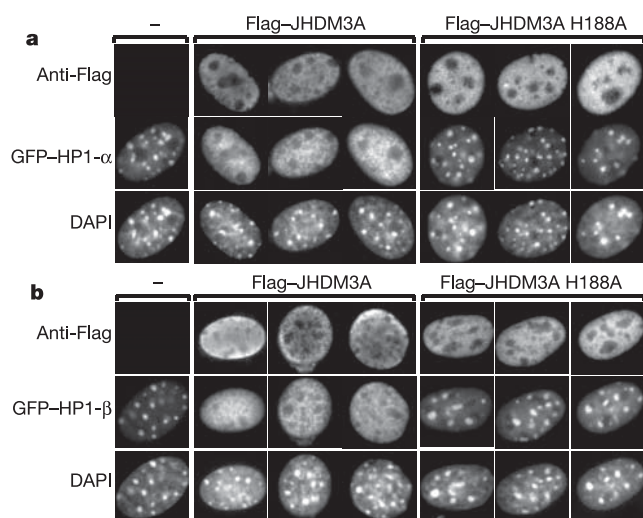


Figure 3 | Overexpression of JHDM3A antagonizes HP1 recruitment to pericentric heterochromatin. **a, b**, Green fluorescent protein (GFP)-tagged HP1- α (GFP-HP1- α) and - β (GFP-HP1- β) show punctate fluorescence corresponding to H3K9me3-containing pericentric heterochromatin in mouse cells (left panels). Overexpression of JHDM3A causes a diffuse re-distribution of GFP-HP1 within the nucleus (middle panels). An intact JmjC domain is essential for antagonizing HP1 recruitment to pericentric heterochromatin, as a mutation in the predicted Fe(II)-binding site abrogated the effect of JHDM3A on GFP-HP1 localization (right panels).

The recent identification of enzymes that can demethylate histone lysine residues has clearly demonstrated the reversibility of this epigenetic modification^{9–11}, and our characterization of JHDM3A demonstrates that trimethyl-lysine is also a reversible modification. Little is known about the function of H3K36 methylation in mammals, but this modification is associated with transcribed regions of active genes¹⁸. A significant amount of H3K9 methylation resides in heterochromatic regions^{21,22}, but this modification also regulates genes found in euchromatic regions^{24,25}. The biological rationale in combining the removal of H3K9 and H3K36 methylation through the activity of one enzyme remains unclear, but suggests that there may be a previously unrealized link between modification of H3K9 and H3K36. Surprisingly, removal of H3K9/H3K36 methylation by JHDM3A is associated with transcriptional repression, indicating that JHDM3A may function to antagonize H3K9/H3K36 methylation, which is associated with gene activity^{12,18}.

Recent biochemical studies have defined a role for the tandem Tudor domain of JHDM3A in the recognition of methyl-lysine residues in histones H3 and H4 (refs 26, 27). It is tempting to speculate that the enzymatic activity of JHDM3A is recruited by the tandem Tudor domain of JHDM3A to chromatin containing specific histone modifications. Further structural and functional analyses of JHDM3A will be instrumental in defining how this interesting demethylase specifically recognizes, demethylates and contributes to the regulation of the H3K9/H3K36 methylation states *in vivo*.

METHODS

Demethylation assay and mass spectrometry. All histone substrates were radioactively labelled as described previously^{10,28}. Equal counts of labelled substrate were used in histone demethylation reactions containing ascorbate. Mass spectrometry was performed also as previously described¹⁰.

Transfection, immunofluorescence microscopy, and western blotting. NIH3T3 and HEK293T cells were grown in DMEM containing 10% FBS and penicillin/streptomycin. For western blot analysis, 10-cm plates were transfected with 20 μ g of expression plasmid using Lipofectamine 2000 reagent (Invitrogen). Cells were harvested 48 h post-transfection, and histones were acid-extracted and

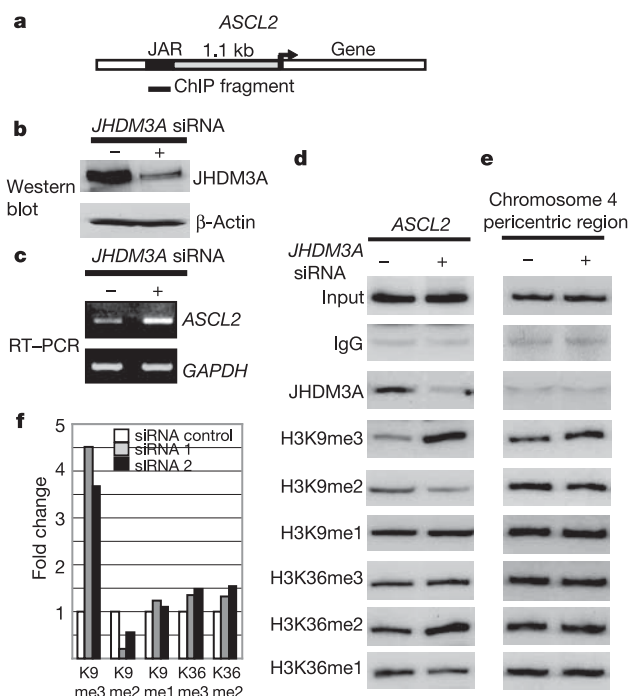


Figure 4 | JHDM3A regulates H3K9/H3K36 methylation at a euchromatic target gene. **a**, A diagram of the human *ASCL2* gene structure. “JAR” represents the ‘JHDM3A-associated region’. **b**, JHDM3A protein levels were efficiently reduced by treatment of cells with *JHDM3A* siRNA. **c**, siRNA-mediated knockdown of JHDM3A results in increased expression of the *ASCL2* gene. **d, e**, Untreated (–) and *JHDM3A*-siRNA-treated (+) cells were used in ChIP assays to analyse histone modifications at the *ASCL2* gene and the chromosome 4 pericentric region. Reduced levels of JHDM3A caused reduced occupancy of JHDM3A at the *ASCL2* gene and increased levels of H3K9me3 and H3K36me2, as assessed by radioactive PCR. JHDM3A was not specifically enriched at the chromosome 4 pericentric regions, and had no effect on the H3K9/H3K36 methylation profiles at this location. **f**, Quantitative real-time PCR analysis of histone modification levels at the *ASCL2* JAR after treatment with two independent siRNAs (see Methods) revealed a increase in H3K9me3 levels and subtle changes in lower H3K9 and H3K36 modifications.

processed for western blot analysis using standard western blotting techniques and enhanced chemiluminescence (ECL) detection. For immunofluorescence, cells grown on coverslips in 6-well plates were transfected with 2–6 μ g of Flag-JHDM3A expression plasmid using Fugene 6 transfection reagent (Roche). In experiments using GFP-HP1- α or - β , 250 ng of expression vector was included in the transfection.

JHDM3A siRNA, RT-PCR analysis and ChIP. siRNA-mediated JHDM3A knockdown, RT-PCR analysis of *ASCL2*, and ChIP analysis were carried out as described¹⁷. Two independent siRNAs targeting *JHDM3A* were used in the experiments (siRNA1: 5'-AAGUUGAGGAUGGUCUUACCU-3'; siRNA2: 5'-AACACAGUUAUUGACCAUACU-3'). A detailed description of this and other relevant methods can be found in Supplementary Information.

Received 16 January; accepted 4 May 2006.

Published online 28 May 2006.

- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
- Flanagan, J. F. et al. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181–1185 (2005).
- Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
- Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).

7. Bannister, A. J., Schneider, R. & Kouzarides, T. Histone methylation: dynamic or static? *Cell* **109**, 801–806 (2002).
8. Duerre, J. A. & Lee, C. T. *In vivo* methylation and turnover of rat brain histones. *J. Neurochem.* **23**, 541–547 (1974).
9. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
10. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).
11. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**, 483–495 (2006).
12. Vakoc, C. R., Mandat, S. A., Olenchok, B. A. & Blobel, G. A. Histone H3 lysine 9 methylation and HP1 γ are associated with transcription elongation through mammalian chromatin. *Mol. Cell* **19**, 381–391 (2005).
13. Clissold, P. M. & Ponting, C. P. JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A $_2$ β . *Trends Biochem. Sci.* **26**, 7–9 (2001).
14. Treweek, S. C., McLaughlin, P. J. & Allshire, R. C. Methylation: lost in hydroxylation? *EMBO Rep.* **6**, 315–320 (2005).
15. Katoh, M. & Katoh, M. Identification and characterization of JMJD2 family genes *in silico*. *Int. J. Oncol.* **24**, 1623–1628 (2004).
16. Gray, S. G. *et al.* Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein. *J. Biol. Chem.* **280**, 28507–28518 (2005).
17. Zhang, D., Yoon, H. G. & Wong, J. JMJD2A is a novel N-CoR-interacting protein and is involved in repression of the human transcription factor achaete scute-like homologue 2 (ASCL2/Hash2). *Mol. Cell. Biol.* **25**, 6404–6414 (2005).
18. Bannister, A. J. *et al.* Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* **280**, 17732–17736 (2005).
19. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
20. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
21. Peters, A. H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
22. Rice, J. C. *et al.* Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol. Cell* **12**, 1591–1598 (2003).
23. Whetstone, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467–481 (2006).
24. Sarraf, S. A. & Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15**, 595–605 (2004).
25. Schultz, D., Ayyanathan, K., Negorev, D., Maul, G. & Rauscher, F. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
26. Kim, J. *et al.* Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* **7**, 397–403 (2006).
27. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. & Xu, R. M. Recognition of histone H3 lysine-4 methylation by the double Tudor domain of JMJD2A. *Science* **312**, 748–751 (2006).
28. Zhang, X. *et al.* Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Fang, B. Strahl, T. Jenuwein, L. Schmiedeberg, A. Verreault and Y. Shinkai for plasmids; X. Cheng and R. Cao for Dim5 protein and EZH2 complex, respectively; L. Lacomis for help with mass spectrometry; and technical assistance from C. Toumazou. This work was supported by NIH grants to Y.Z., P.T. and J.W. Y.Z. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions R.J.K. carried out most of the experiments in Figs 1–3 and the Supplementary Figures; K.Y. generated recombinant protein; H.E.-B. and P.T. performed mass spectrometric analysis; Y.B., D.Z. and J.W. carried out the experiments in Fig. 4; R.J.K. and Y.Z. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.Z. (yi_zhang@med.unc.edu).

Atom-by-atom analysis of global downhill protein folding

Mourad Sadqi¹, David Fushman¹ & Victor Muñoz¹

Protein folding is an inherently complex process involving coordination of the intricate networks of weak interactions that stabilize native three-dimensional structures. In the conventional paradigm, simple protein structures are assumed to fold in an all-or-none process¹ that is inaccessible to experiment. Existing experimental methods therefore probe folding mechanisms indirectly. A widely used approach interprets changes in protein stability² and/or folding kinetics^{3,4}, induced by engineered mutations, in terms of the structure of the native protein. In addition to limitations in connecting energetics with structure⁵, mutational methods have significant experimental uncertainties⁶ and are unable to map complex networks of interactions. In contrast, analytical theory predicts small barriers to folding and the possibility of downhill folding^{7,8}. These theoretical predictions have been confirmed experimentally in recent years^{9–11}, including the observation of global downhill folding¹². However, a key remaining question is whether downhill folding can indeed lead to the high-resolution analysis of protein folding processes¹³. Here we show, with the use of nuclear magnetic resonance (NMR), that the downhill protein BBL from *Escherichia coli* unfolds atom by atom starting from a defined three-dimensional structure. Thermal unfolding data on 158 backbone and side-chain protons out of a total of 204 provide a detailed view of the structural events during folding. This view confirms the statistical nature of folding, and exposes the interplay between hydrogen bonding, hydrophobic forces, backbone conformation and side-chain entropy. From the data we also obtain a map of the interaction network in this protein, which reveals the source of folding cooperativity. Our approach can be extended to other proteins with marginal barriers (less than $3RT$), providing a new tool for the study of protein folding.

According to a variety of thermodynamic tests, the 40-residue domain BBL folds in a globally downhill fashion^{12,14,15}. Global downhill proteins should fold into defined three-dimensional (3D) structures under native conditions, and should unfold in a structurally continuous process as destabilization increases¹³. To test this idea, we investigated the thermal unfolding of our 40-residue variant of BBL (henceforth referred to as Naf-BBL) by two-dimensional proton NMR. We performed all NMR experiments in Naf-BBL at pH 5.3 to facilitate comparison with previous NMR structural studies¹⁶ on a longer BBL variant that has higher stability¹⁷. The structural analysis of Naf-BBL was performed at 278 K because at this temperature the thermodynamic stability and the H_α chemical shifts of Naf-BBL match very closely those of the longer BBL variant at the 298 K used in previous structural studies (Supplementary Information). The backbone superimposition of the 20 lowest-energy NMR structures for Naf-BBL is shown in Fig. 1a. The ensemble of structures has an average root-mean-square deviation (r.m.s.d.) of 0.48 Å for backbone atoms of non-tail residues (5–38), and an r.m.s.d. of 1.1 for all heavy atoms of the same residues (Supplementary Table S1), indicating that the 3D

structure is well defined. The structures of the two variants are essentially identical, as shown in the superimposition of Fig. 1b. These results illustrate that the native ensemble of the BBL domain is independent of the length of its tails and has the structural features of a regular folded globular protein. ^{15}N -NMR relaxation measurements in Naf-BBL chemically synthesized with 16 ^{15}N isotopic labels (all Gly, Val, Ala and 5 Leu residues) also indicate that the 278 K native ensemble corresponds to a defined structure without large-amplitude motions in the subnanosecond and nanosecond time-scales (Fig. 1c–e). A model-free analysis produces S^2 values higher than 0.8 for all labelled residues in the structured segment of the protein (namely residues 5–38) (Supplementary Fig. S3). However, the ^{15}N R_2 values of several residues were higher than predicted from the model-free assumption (bars above the dashed line in Fig. 1d), signalling the presence of exchange processes on the submillisecond timescale. This is expected because at 278 K Naf-BBL is at the edge of the unfolding pretransition region (Fig. 2c). At 278 K, ^{15}N transverse relaxation dispersion experiments produced characteristic dispersion profiles for residues with increased R_2 and flat profiles for residues with no signs of exchange in R_2 (Fig. 1f, g). Quantitative analysis indicated that all observed exchange processes occur with a single rate of $1/(150\ \mu\text{s})$ (Supplementary Information). This rate is identical to the global folding dynamics measured at this temperature with nanosecond T -jump experiments (F. Y. Oliva and V.M., unpublished observations). Relaxation dispersion experiments at 293 K produced flat profiles for all labelled residues (Fig. 1f, g), which were all consistent with a single exchange rate equal to the $1/(30\ \mu\text{s})$ measured with the T -jump technique. It is obvious from these results that Naf-BBL is in the fast exchange regime even at the lowest temperatures.

We then investigated the thermal unfolding of Naf-BBL by measuring proton chemical shifts as a function of temperature. Because Naf-BBL is in the fast exchange regime, observed chemical shifts correspond to ensemble averages. Out of 204 sets of assigned protons, we identified 158 (36 H_N , 40 H_α and 82 side-chain protons) that could be resolved in the 268–338 K range and move towards characteristic random-coil values at increasing temperature (Supplementary Fig. S4 and Supplementary Table S2). The 158 atomic unfolding curves have a rich pattern of behaviours, with two-thirds being sigmoidal and the rest showing two apparent transitions (Fig. 2a). The 158 curves were analysed with phenomenological two-state or three-state models (Methods and Supplementary Information). This analysis produced a broad range of midpoint temperatures (T_m) spanning about 60 K. The T_m distribution is approximately gaussian, with a mean of about 304 K and a standard deviation of about 17 K (Fig. 2b). The same mean T_m and a slightly broader distribution were obtained directly from the derivative of the unfolding curves, a procedure that does not depend on pre-transition and post-transition baselines (data not shown). In spite of such complexity,

¹Department of Chemistry and Biochemistry, and Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, USA.

the average atomic unfolding behaviour obtained by singular value decomposition of the chemical shift versus temperature matrix produces a smooth sigmoidal curve (Fig. 2c). Furthermore, the average NMR transition is very similar to the circular dichroism unfolding curve, including 'apparent' T_m within 4 K of each other (inset of Fig. 2c). The sum of atomic unfolding processes obtained by NMR therefore corresponds to the global process monitored by a low-resolution local technique. These results reveal the statistical nature of protein folding. The unfolding process of Naf-BBL is characterized by partly overlaid atomic unfolding events spanning the complete global transition instead of the linear combination between folded and unfolded molecules that is conventionally assumed. The roughly gaussian distribution implies that many atomic unfolding events are similar to the mean and thus to each other (Fig. 2d). Finding a subset of signals with mean-like behaviour is therefore always possible and should not be construed as evidence for two-state folding. The degree of correlation between atomic unfolding events at each temperature can be observed in histograms of the degree of nativeness (Supplementary Fig. S6). At the global midpoint (304 K), the histogram shows an almost uniform distribution with atoms experiencing environments ranging from all native to all unfolded (Fig. 2e). The distribution is maximally broad at the midpoint and narrows down at lower and higher temperatures (Fig. 2f). These observations show that Naf-BBL unfolds in a structurally continuous transition in which the midpoint corresponds to the largest structural heterogeneity. In other words, the folding free-energy surface has a single well (minimum) that gradually shifts from native to unfolded at increasing temperature¹³. Conformational fluctuations in a single well are diffusive and have timescales simply determined by the diffusion coefficient and the overall shape of the well¹⁸, which explains the observation of similar exchange rates for all residues in the NMR transverse relaxation dispersion experiments.

A cluster analysis of the 158 atomic unfolding curves shows that groups of protons with similar T_m values do not localize in specific

structural regions (Fig. 3a). The continuous unfolding of Naf-BBL is therefore not characterized by a sequence of local unfolding events (that is, a hierarchical assembly of modules). The protein does not have a unique folding mechanism; rather, it has families of mechanisms that must be described in statistical terms. This situation contrasts with the stepwise unfolding observed by hydrogen-exchange experiments on cytochrome *c* (ref. 19). However, clustering based on proton chemical properties reveals the interplay between the forces driving folding. In absence of ring-current effects (Naf-BBL lacks aromatic residues in the structured region), H_N chemical shifts mostly sense the hydrogen-bonding status of the peptide bond²⁰ (Fig. 3b), H_α chemical shifts are primarily sensitive to flanking ϕ and ψ dihedral angles²¹ (Fig. 3c), and hydrophobic side-chain protons report on the tertiary structure (Fig. 3d). Figure 3b–d reveals that a large fraction of hydrogen bonds break at very low temperatures (first lobe with $\langle T_m \rangle \approx 289$ K in Fig. 3b). Melting of tertiary structure occurs around the average midpoint temperature (about 304 K, atoms in the crevice between the two helices in Fig. 3d), accompanied with the largest fraction of global unfolding. Backbone conformation unfolds at the highest temperatures ($\langle T_m \rangle \approx 309$ K; Fig. 3c). At this point, the second group of hydrogen bonds, which locate mainly in the structured loop (residues 20–30), also break (second lobe with $\langle T_m \rangle \approx 311$ K in Fig. 3b). This loop seems to be autonomously stabilized by a dense network of local contacts (see red contacts in Fig. 4b). These results imply that there is little coupling between tertiary interactions and backbone structure. The entropy penalty of fixing tertiary interactions seems to be mostly paid by the side chains, whereas backbone conformation is stabilized by local contacts and intrinsic propensities. In fact, H_α chemical shifts indicate the presence of native-like backbone conformation even in the thermally unfolded state. The available kinetic data on two-state and three-state proteins are fully consistent with this sequence of events. The low- T_m hydrogen bonds are equivalent to fast-exchanging protons in hydrogen-exchange experiments. The breaking of tertiary structure as the major unfolding event is consistent with recent

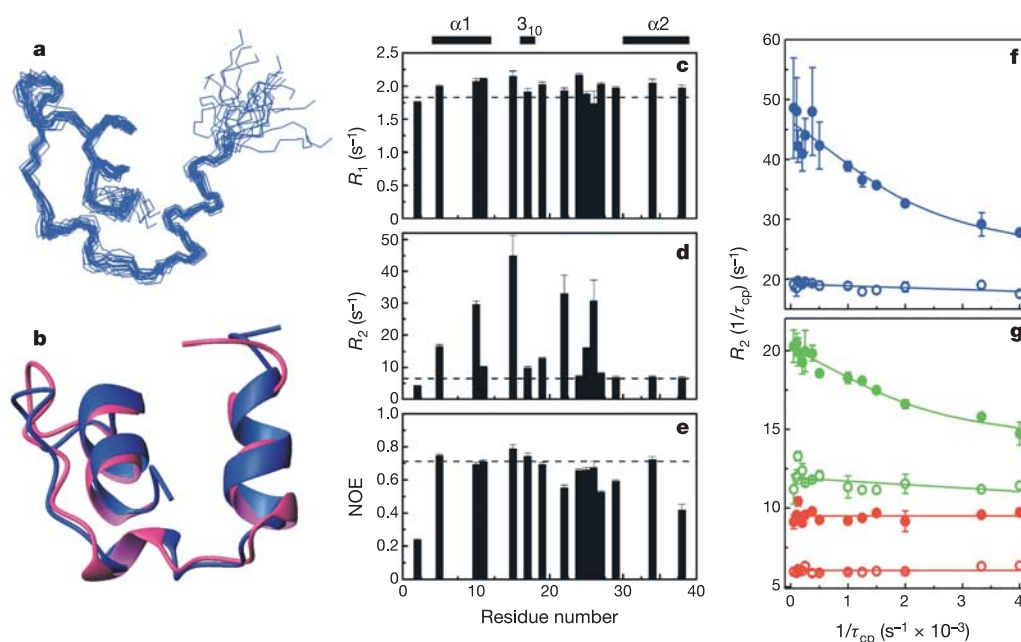


Figure 1 | Structure and dynamics of the native ensemble. **a**, Backbone superimposition of the ensemble of 20 lowest-energy structures of Naf-BBL calculated from NMR data at pH 5.3 and 278 K. **b**, Superimposition of the lowest-energy structure of Naf-BBL (blue) and the average minimized structure from ref. 16 (magenta). **c**, Longitudinal ^{15}N relaxation rate R_1 . **d**, Transverse ^{15}N relaxation rate R_2 . **e**, Heteronuclear Overhauser effect (NOE). **f**, **g**, ^{15}N relaxation dispersion curves (symbols) and fits²⁸ (lines) for

Gly 22 (**f**), Val 25 (**g**, green) and Val 34 (**g**, red) at 278 K (filled circles) and 293 K (open circles). A single exchange rate of $1/(150 \mu\text{s})$ at 278 K and of $1/(30 \mu\text{s})$ at 293 K was used for the fits of all signals in the data set. The error bars in **c–e** represent uncertainties in the parameters (one standard deviation) given the experimental noise in the NMR spectra. The error bars in **f** and **g** correspond to the square root of the variance (standard deviation) for three independent ^{15}N relaxation dispersion measurements.

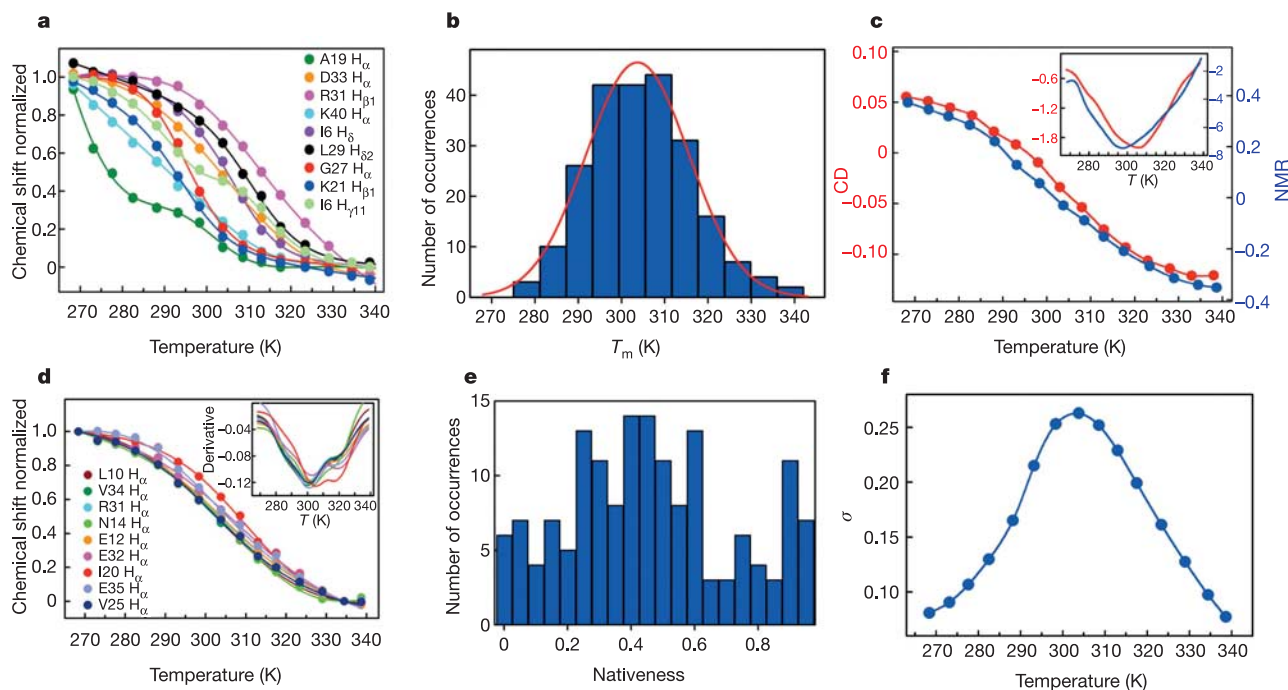


Figure 2 | Protein thermal unfolding atom by atom. **a**, Plot of chemical shifts against temperature for nine representative protons. **b**, Histogram of T_m values for all 158 sets of protons. Protons displaying three-state behaviour (for example green curves in **a**) provide two T_m values to the histogram. $\langle T_m \rangle = 303.7$ K; $\sigma_{T_m} = 16.9$. **c**, Comparison between low-resolution thermal unfolding (circular dichroism, red) and the average of the 158 normalized atomic unfolding curves (NMR, blue). The y-axis represents the amplitude of the second singular value for the circular

dichroism spectra versus T (red), or for the matrix of 158 NMR chemical shifts versus T (blue). The inset shows the derivative of the curves. **d**, Plot of chemical shifts against temperature for a set of 9 H_α displaying mean-like behaviour. **e**, Histogram of the 158 native probabilities at 303.8 K obtained from individual two-state or three-state fits. **f**, Standard deviation for the 158 native probabilities as function of temperature (histograms are shown in Supplementary Fig. S6).

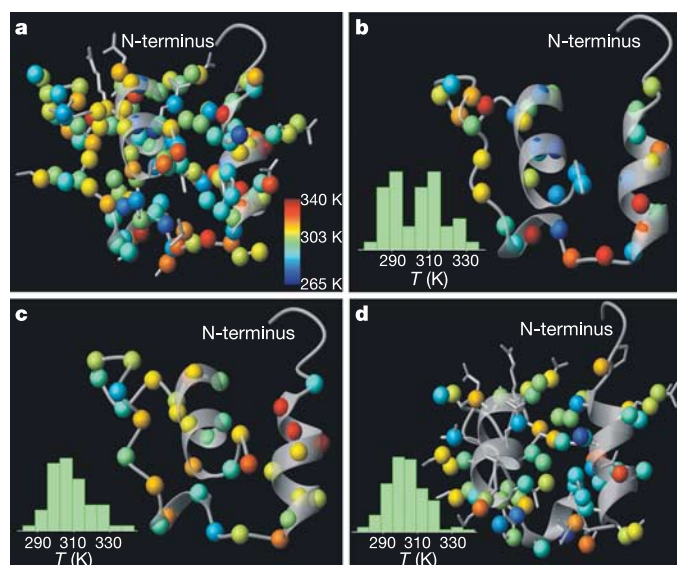


Figure 3 | Mapping thermal stability on the 3D structure. **a**, Diagram of Naf-BBL structure showing the thermal stability of the 158 protons mapped on their corresponding heavy atoms. The colour displayed in each heavy atom corresponds to the average T_m for all transitions (that is, 1 for two-state and 2 for three-state) of all protons (for example, 2 H_α for glycine residues) that are mapped to that heavy atom. **b**, As **a** for only amide protons. **c**, As **a** for α -protons. **d**, As **a** for side-chain protons. The insets in **b–d** show the T_m histograms.

descriptions of the transition state for two-state folding¹¹ and with the widespread observation of intermediate ϕ values⁶. Finally, the 310-K molten globule-like stage in Naf-BBL (native-like $\phi - \psi$ distribution, almost no tertiary interactions, and highly compact¹²) has common features with the burst-phase detected in stopped-flow folding experiments²². These similarities indicate a common overall folding process for all globular proteins, regardless of the presence or absence of folding barriers.

Our results raise an interesting question related to folding cooperativity. Naf-BBL seems to defy the commonly held notion that folding must be cooperative if a protein is to fold into a 3D structure. However, the key question is really a quantitative one: how much thermodynamic coupling is required for a protein to fold into a defined globular structure? To address this question we introduce an index of folding cooperativity based on the similarity among atomic unfolding transitions, which we term the mean thermodynamic coupling index (MTCI; Methods and Supplementary Information). MTCI is a simple measure of the similarity (in r.m.s.d. terms) between all possible pairs of normalized atomic unfolding transitions in the protein. The MTCI tends to infinity for an all-or-none transition (that is, infinite coupling), and to zero for a set of totally uncoupled unfolding transitions. For the 158 Naf-BBL protons we obtain $MTCI \approx 1.2$ (in RT units). Thus, although Naf-BBL unfolding is very far from an all-or-none transition, it displays a measurable degree of cooperativity. Such weak thermodynamic coupling is enough to fold into a 3D structure under native conditions, but at increasing temperatures the atoms (now with higher thermal energy) progressively decouple from each other. The MTCI provides a general scale for quantifying folding cooperativity. Computing this parameter from molecular simulations of unfolding (for example, by the replica-exchange method) is straightforward, providing a tool with which to compare experiment and simulation directly. The same

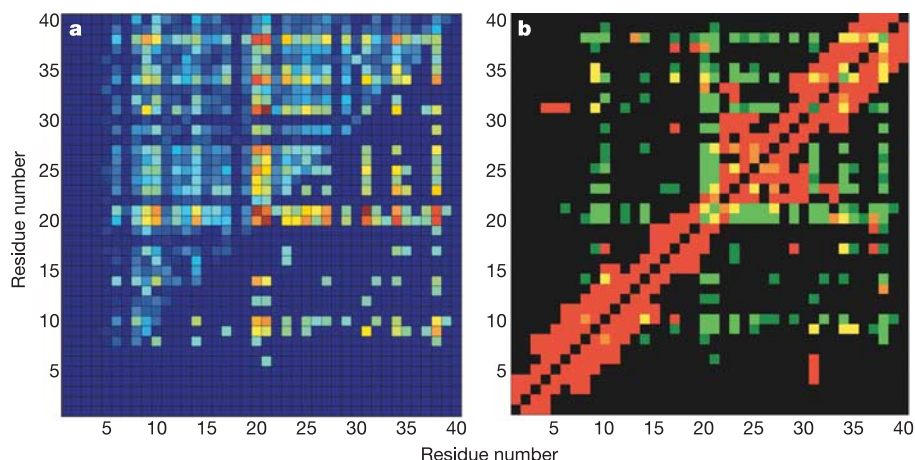


Figure 4 | Interaction matrices and the emergence of folding cooperativity. **a**, Matrix of thermodynamic couplings for Naf-BBL obtained with equation (2) in Methods. Dark red signifies the strongest coupling (maximum is 22) and dark blue the weakest (minimum is -10). The upper left triangle shows all the data; the lower right triangle shows

couplings with at least one-third of the intensity of the strongest coupling. **b**, Superimposition of the matrix of strongest couplings (lower triangle of **a**) binned in two levels for simplicity (green, with brighter meaning stronger), and the matrix of proton-proton distances less than 0.5 nm as obtained from the 3D structure (red).

approach can be taken further to estimate a matrix of pairwise residue-residue interactions (Fig. 4a, Methods and Supplementary Information). Figure 4a shows that the sparse strong couplings (see lower right triangle) are distributed in a complex 3D network. Figure 4b shows a superimposition of the matrix of strong couplings (shown in green) and the contact map based on proton-proton distances less than 0.5 nm (shown in red). Yellow and orange dots in this graph signify strong couplings between residues in direct spatial contact. The graph exposes a subset of tertiary contacts that are critical for the stabilization of the native structure: contacts of Ile 20 with Val 34 and His 37; Leu 9 with Arg 31 and Val 34; Ala 17 with Val 34; and His 13 with Leu 38. There also are clear local couplings in the loop region (residues 21–28), and local couplings extending through the second α -helix (residues 31–38). Most of the first α -helix, which is intrinsically stable, displays little coupling. The pattern of green dots shows coupled residues that are not in spatial contact. Most of these couplings seem to propagate through secondary contacts (that is, two residues connected through a third residue) or even higher-order contacts. For example, Ile 20 couples to most of the carboxy-terminal half through only three primary contacts (Fig. 4b). Second-order and higher-order couplings suggest the presence of many-body interactions. The existence of many-body terms in protein folding has been proposed theoretically²³ and implemented recently in off-lattice models²⁴. Our results indicate that folding cooperativity emerges from these higher-order couplings.

In principle, the approach described here can be generalized to all proteins with folding barriers smaller than $3RT$ (about 5% of the population at the top relative to the minimum). Recent analysis of size-scaling of folding times points to proteins of less than 60 residues as potential targets²⁵; α -helical proteins in general, and coiled coils in particular, seem promising candidates²⁶. Our approach might even work for some proteins previously categorized as proteins that undergo two-state folding, as suggested in recent theoretical studies²⁷.

METHODS

The production and handling of non-labelled BBL samples were performed as described previously^{12,17}. Naf-BBL with ^{15}N labels in residues 2, 5, 10, 11, 15, 17, 19, 22, 24, 25, 26, 27, 29, 34, 38 and 39 was chemically synthesized by Biopeptide Co. Standard ^1H NMR experiments were performed in the range 268–338 K using a DRX-500 MHz Bruker spectrometer as described previously¹². The procedures for NMR structure calculations are described in Supplementary Information. The rates of longitudinal (R_1) and transverse (R_2) ^{15}N relaxation and ^1H - ^{15}N NOEs were measured at 600 MHz (^1H frequency) using standard

methods. ^{15}N transverse relaxation dispersion measurements were performed using the method of ref. 29. Effective relaxation rates (R_2^{eff}) were obtained by comparing the signals in a reference spectrum (no GPMG) and in a series of 14 two-dimensional ^1H - ^{15}N spectra recorded with the same total CPMG period of 80 ms, and with the time interval between the 180° CPMG pulses ranging from 250 μs to 20 ms. Chemical shifts at different temperatures were referenced to 2,2-dimethyl-2-silapentane-3-sulphonate (DSS). Individual fits of proton chemical shifts as a function of temperature were performed with two- or three-state models and minimal number of adjustable parameters (that is, ignoring changes in heat capacity). The temperature dependence of the fitted baselines was constrained within narrow, physically reasonable ranges (± 0.015 p.p.m. K^{-1} for H_α and side-chain protons, and ± 0.02 p.p.m. K^{-1} of the typical values of unstructured peptides for H_N). More details are given in Supplementary Information. The MTCI introduced here is defined as:

$$\text{MTCI} = \ln \left(\frac{\left\langle \sqrt{(u_i - u_j)(u_i - u_j)^T} \right\rangle}{\left\langle \sqrt{\delta((p_i - p_j)(p_i - p_j)^T - (n_i - n_j)(n_i - n_j)^T)} \right\rangle} \right) \quad (1)$$

where p_i and p_j are row vectors from a matrix with the native probability as a function of temperature for m different probes (obtained in individual thermodynamic fits), n_i and n_j are row vectors from a matrix of the fluctuations in the mean native probability arising from experimental uncertainty, and u_i and u_j are row vectors from a matrix of the native probability for m random unfolding transitions spanning the global unfolding transition. δ is 0 for $i = j$ and 1 otherwise. 158 n_i vectors were generated by adding statistical noise of magnitude equivalent to the sum of least squares of each individual fit to a theoretical unfolding curve produced with the global unfolding parameters. 158 u_i vectors were generated by calculating theoretical unfolding curves with thermodynamic parameters selected randomly within a range consistent with the global unfolding transition. The denominator in equation (1) corresponds to the r.m.s.d. between all possible pairs of atomic thermal unfolding curves ($m(m-1)/2$ for m atoms) corrected for the differences in r.m.s.d. arising from experimental uncertainty. The numerator normalizes the MTCI value to the broadness of the global unfolding transition. This correction facilitates comparison between proteins of different size and/or different apparent T_m (that is, the higher the T_m or size, the sharper the unfolding transition). Pairwise residue interactions are estimated from the couplings between all the atoms of one residue with all the atoms of another as:

$$\text{TCI}_{xy} = \sum_i \sum_j \ln \left(\frac{\left\langle \sqrt{\delta(p_m - p_n)(p_m - p_n)^T} \right\rangle}{\sqrt{(p_{x,i} - p_{y,j})^2}} \right) \quad (2)$$

where i runs over all atoms of residue x and j over all atoms of residue y , and the numerator corresponds to the mean r.m.s.d. for all atomic curves in the protein. The index is positive when the sum of atomic couplings for two residues is stronger than the mean protein coupling, and negative when is weaker.

Received 6 February; accepted 2 May 2006.

Published online 14 June 2006.

1. Tanford, C. Protein denaturation. *Adv. Prot. Chem.* **23**, 121–282 (1968).
2. Matthews, B. W. Structural and genetic-analysis of protein-folding and stability. *Curr. Opin. Struct. Biol.* **3**, 589–593 (1993).
3. Goldenberg, D. P. in *Protein Folding* (ed. Cheighton, T. E.) 353–406 (Freeman, New York, 1992).
4. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782 (1992).
5. Honig, B. & Yang, A. S. Free-energy balance in protein-folding. *Adv. Protein Chem.* **46**, 27–58 (1995).
6. Sanchez, I. E. & Kiefhaber, T. Origin of unusual ϕ -values in protein folding: Evidence against specific nucleation sites. *J. Mol. Biol.* **334**, 1077–1085 (2003).
7. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein-folding—a synthesis. *Prot. Struct. Funct. Genet.* **21**, 167–195 (1995).
8. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
9. Yang, W. Y. & Gruebele, M. Folding at the speed limit. *Nature* **423**, 193–197 (2003).
10. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding ‘speed limit’. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).
11. Akmal, A. & Muñoz, V. The nature of the free energy barriers to two-state folding. *Prot. Struct. Funct. Bioinformatics* **57**, 142–152 (2004).
12. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M. & Muñoz, V. Experimental identification of downhill protein folding. *Science* **298**, 2191–2195 (2002).
13. Muñoz, V. Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. *Int. J. Quantum Chem.* **90**, 1522–1528 (2002).
14. Oliva, F. Y. & Muñoz, V. A simple thermodynamic test to discriminate between two-state and downhill folding. *J. Am. Chem. Soc.* **126**, 8596–8597 (2004).
15. Muñoz, V. & Sanchez-Ruiz, J. M. Exploring protein folding ensembles: A variable barrier model for the analysis of equilibrium unfolding experiments. *Proc. Natl Acad. Sci. USA* **101**, 17646–17651 (2004).
16. Robien, M. A. *et al.* 3-Dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyl transferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. *Biochemistry* **31**, 3463–3471 (1992).
17. Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Muñoz, V. Robustness of downhill folding: Guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* **44**, 7435–7449 (2005).
18. Hanggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).
19. Maity, H., Maity, M., Krishna, M. M. G., Mayne, L. & Englander, S. W. Protein folding: The stepwise assembly of foldon units. *Proc. Natl Acad. Sci. USA* **102**, 4741–4746 (2005).
20. Wagner, G., Pardi, A. & Wüthrich, K. Hydrogen-bond length and H-1-NMR chemical-shifts in proteins. *J. Am. Chem. Soc.* **105**, 5948–5949 (1983).
21. Wishart, D. S. & Sykes, B. D. Chemical-shifts as a tool for structure determination. *Nucl. Magn. Res. C* **239**, 363–392 (1994).
22. Arai, M. & Kuwajima, K. Rapid formation of a molten globule intermediate in refolding of α -lactalbumin. *Fold. Des.* **1**, 275–287 (1996).
23. Plotkin, S. S., Wang, J. & Wolynes, P. G. Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J. Chem. Phys.* **106**, 2932–2948 (1997).
24. Shimizu, S. & Chan, H. S. Anti-cooperativity and cooperativity in hydrophobic interactions: Three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. *Prot. Struct. Funct. Genet.* **48**, 15–30 (2002).
25. Naganathan, A. N. & Muñoz, V. Scaling of folding times with protein size. *J. Am. Chem. Soc.* **127**, 480–481 (2005).
26. Holtzer, M. E., Lovett, E. G., d’Avignon, D. A. & Holtzer, A. Thermal unfolding in a GCN4-like leucine zipper: C-13 α NMR chemical shifts and local unfolding curves. *Biophys. J.* **73**, 1031–1041 (1997).
27. Klimov, D. K. & Thirumalai, D. Is there a unique melting temperature for two-state proteins? *J. Comp. Chem.* **23**, 161–165 (2002).
28. Palmer, A. G., Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238 (2001).
29. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states on an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. de Alba for help and guidance in the calculation of NMR structures with Xplor-NIH and for generating the table of structure statistics; P. Vasos for assistance with the NMR relaxation experiments; and Y. Hathout for mass-spectrometric analysis. The research described in this article was supported by the NIH and the NSF.

Author Contributions D.F. performed and analysed the NMR relaxation experiments and contributed to manuscript preparation. M.S. performed all other NMR experiments and circular dichroism experiments, assigned all the NMR spectra, and contributed to manuscript preparation. V.M. conceived the project, supervised the work, developed the methods for analysis, analysed the experimental data, and contributed to manuscript preparation. The structural analysis was performed in conjunction by M.S. and V.M.

Author Information The atomic coordinates of Naf-BBL have been deposited in the Protein Data Bank with the accession number 2QYU. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.M. (vmunoz@umd.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

On the face of it, it sounds like a failure for Scotland. A Scottish biotech company that last year was bolstered by government funds to the tune of £5 million (US\$9 million) has recently merged with a US company and moved to the other side of the Atlantic. Why invest so much in a local company, only for it to move abroad? But Cyclacel's pond-hopping could be viewed as a success of sorts, and a harbinger of things to come.

Most of the company's research and development (R&D) staff will stay in Scotland. And having a corporate presence in the United States means better access to US investment dollars. After the merger in March with Xcyte Therapies of Seattle, Washington, the new company used Xcyte's stock listing to raise \$45 million.

Cyclacel is not alone. Last August, French drug company IDM merged with Epimmune in San Diego, California, again gaining access to Nasdaq — and IDM kept most of its R&D employees in France. And in May, after a merger with CancerVax, German company Micromet gained a Nasdaq listing and established a corporate base in Carlsbad, California, while retaining most of its employees in Germany.

This strategy may help explain how, in some respects, European biotech companies are beating their US counterparts — at least in terms of sheer number. Ten years ago, Europe had 584 biotech companies and the United States had 1,308. Now Europe boasts 1,613 biotech firms, whereas the United States has 1,415.

Where the money comes from is not really an issue for working scientists — they are more interested in whether or not they will have better job opportunities. If the net number of biotech jobs grows, then workers are pretty unlikely to care where their corporate office is located. And even if their jobs eventually become outsourced — as has been the case with some large global pharmaceutical firms — European biotech workers can always do what their companies did: follow the money.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
Natureevents: Matt Betteridge (4015)
UK/RoW/Ireland:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Claudia Paulsen Young (4973)
France/Switzerland/Belgium/Italy:
Maria Ceraolo (4994)

Germany/Austria/The Netherlands:
Reya Silao (4970)
Business Development Manager:
Amelie Pequignot (4974)

Advertising Production Manager:
Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock
Naturejobs online production:
Catherine Alexander

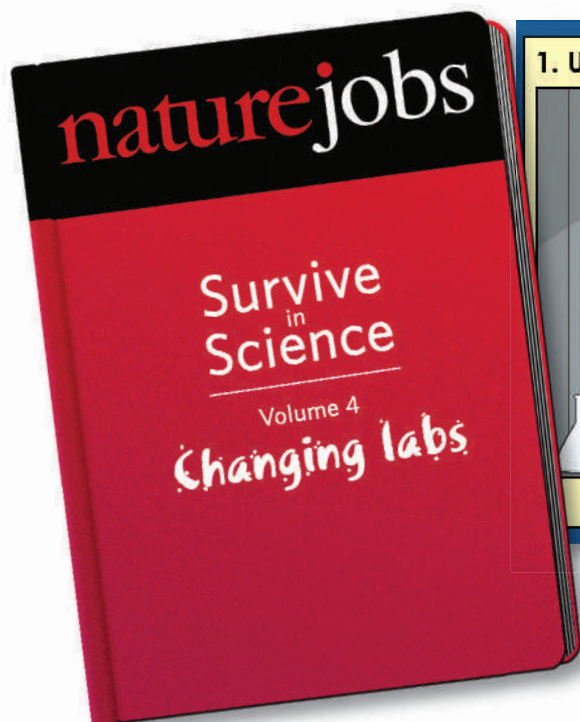
European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

US Head Office, New York
75 Varick Street,
9th Floor, New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:
Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



1. Unhappy? Consider changing labs



2. But look before you leap!

Making a move

On the very first day of Hans Weber's* first doctoral project, his adviser showed him to a freshly renovated room with a host of equipment for electrophysiology experiments that needed to be assembled and installed. "I had the manual in one hand and a screwdriver in the other," Weber recalls.

Two months in and ready to start his first experiment at the research centre in central Germany, he needed to calibrate the amplifier that powered the experiments. It blew up on the first try. "I destroyed a €15,000 amplifier and I thought: 'This should not be the way a PhD goes,'" he says.

It took a year for Weber to realize that a lack of supervision and a near-impossible project did not make him a bad or lazy scientist. But like many before him, he carefully weighed up whether he should persevere or move to another laboratory.

Deciding to move — whether because of an unhappy situation, a change in interests or a lab group relocating — almost certainly means some lost time and a transition to an unknown situation. But most lab-switchers say the benefits of changing groups usually outweigh any negatives.

Should you stay or should you go?

Weber admits that he had days of self-doubt and considered quitting science altogether — a not uncommon reaction to this kind of situation. But he realized when talking to family and friends that the neuroscience fanatic in him was going to prevail. And after two older scientists encouraged him to find another group, he began discreetly e-mailing, calling and eventually visiting other groups around Germany.

"I didn't see any chance that things would change in that first lab," he says. "I think when you doubt your PhD for longer than two or three months, you should just change." Weber eventually joined a neurobiology group at a different university.

Although it seemed like a natural progression to him, Weber weighed the pros and cons of his situation,

Morale, money or moving house can all be reasons for switching labs mid-project.

Kendall Powell learns from those who have made the jump with success.

both in and out of the lab, and he spoke to people at the university and outside it. First, figure out if the problems you are facing come from your project, your lab, your university or your environment.

"If you are enjoying your work, your project, then don't quit just because you don't like your location," advises Mhairi Dupré, a first-year doctoral student in evolutionary biology at the University of Oxford, UK. She left a PhD at a university in Montreal, Canada, and took some time away from science before re-enlisting at Oxford a year later. But, if you are dragging yourself into the lab with dread, says Dupré, "There's no point going on with something you are not enjoying."

Make the smallest move that will accomplish your goals: for example, switching projects in the same lab will be easier than switching labs at the same university, which, in turn, will be easier than changing universities.

Sudarshan Anand, a fourth-year student, decided to switch from a general molecular-biology graduate programme at a Maryland university to an immunology programme at the Mayo Clinic in Rochester, Minnesota. He moved again when his supervisor relocated the lab from Minnesota to Johns Hopkins University in Baltimore, Maryland. Moving between programmes, universities and states was a hassle, but proved worthwhile for his long-term plan of doing translational research.

Students and postdocs say that the biggest question you should ask yourself is what resources you need for your planned career path and which lab best fits that list.

The second half of the decision-making process should probe external perspectives. When Carmen Drahl's adviser moved to a new university, she drew on "contacts I did not even realize I had" to help her decide whether to move with her adviser, stay behind and be advised remotely, or switch labs altogether. She talked to her family, college professors, an industry researcher she met at a seminar, her group colleagues and the faculty members at her graduate department.

"Branch out and use your resources," she says.

"There was not a unanimous position, but I went with my gut and the consensus that another lab might hold new opportunities for me." If your adviser takes a new position, there are a few considerations to take into account before deciding whether to move with them. There are also things advisers can do to make that decision less painful (see 'Trading places').

Staying behind and being advised from afar may seem attractive, but it means less face-to-face time and delays in trouble-shooting. The problems multiply if the supervisor moves into an administrative role or an industry job, where their former research is no longer their primary focus.

"There's a reason why labs have principal investigators," says Rob, a biochemist who moved his lab group across the United States to build a large research programme at another university.

Rob recalls that communication became difficult with lab members who stayed behind, and the project's momentum flagged. It is best if those members move into new labs to get local support, he says. He also notes that moving can be more attractive if the lab head ensures that downtime for experimental work is minimized. "That there is going to be huge down-time is a fallacy," says Rob: his lab members who moved were running experiments in about three weeks. "We waste a lot more time doing the wrong experiments than we did moving."

Now what?

Once you decide to switch between laboratory groups, how do you ensure it's not 'out of the frying pan, into the fire'? When looking for a new lab home, you must put your best foot forward to impress that group. But you must also critically evaluate your potential new lab mates and adviser.

Do your homework on groups that interest you: find out their publication history, assess their resources and equipment, and see how successful former lab members have been. Then begin conversations with

"If you are enjoying your work, your project, then don't quit just because you don't like your location. But there's no point going on with something you are not enjoying."

TRADING PLACES

Moving a lab between universities can be difficult — both physically and emotionally, says Rob, a biochemist who moved his lab recently. He and other lab heads offer these tips to smooth the transition.

- 1 Tell your group as soon as possible. Avoid making a vague announcement that could create rumours or panic. Get lab members' input on your decision.
- 2 Speak with lab members individually about what is best for them. Keep workers informed at each step. If budgets allow, have workers visit the new location.
- 3 Realize that you may lose a substantial portion of your group. Anticipate your re-staffing needs.
- 4 Make arrangements to get your new lab up and running as quickly as possible. If you need major renovation or new equipment, consider delaying the move.
- 5 If you decide to support workers staying behind, set clear timelines for research and salary endpoints. **K.P.**

those groups and determine how you might fit into their research (see *Nature* **422**, 784–785; 2003).

During his visits, Weber says he asked himself a number of questions: "How is the team? How is communication? Do they help each other? Is the supervisor open-minded when you discuss the project you might do or are they strict about it?"

Samantha Zeitlin, a postdoc at the University of California, San Diego, says her priorities changed when she began looking for a new lab. Her former adviser was thinking of retiring and she needed a place to finish her project and publications. "I had to stay local," she says. "I picked my new adviser because she was really open to my project. She also had money and space."

Determined to finish her original project, Zeitlin made it her top priority to find a lab that would support that effort.

Drahl also had a slightly different set of criteria when choosing a graduate adviser the second time around. She looked for a lab with stable funding, successful graduates and teamwork. "The research was definitely a factor, but I was looking more at a cluster of practical needs."

You also have to convince a new adviser to take you. Drahl made sure to point out her assets — a fellowship from the National Science Foundation that would cover some of her salary, and a deeper knowledge of biochemistry that would help in some of the lab's collaborations. Zeitlin adds that you need to take yourself seriously, so that others will too. "Don't wait until you have that look of desperation in your eyes," she says.

If you decide to switch between fields or programmes, be prepared to play a bit of catch-up by reading and doing extra coursework. Drahl sat in on classes to absorb more organic chemistry and relied heavily on the technical guidance from her new senior lab mates.

Added bonuses

Although changing labs may seem like a setback in a young science career, many switchers say the relatively short time lost is usually worth the outcome. Anand, who not only switched graduate programmes, but also later moved with his new adviser to another university, says all the transitions have actually given him "the best of both worlds". He says he has a much broader training in molecular biology, immunology and computational biology than he would have had otherwise.

Their transitions may have made for a rough couple of years, but Weber and Dupré both say that they would do it all over again because they now have exciting projects and a deep sense of ownership of their research. "Sitting at my microscope, I forget about everything else. I could talk about my project for hours," says Weber.

Zeitlin has followed her chromosome-biology projects through many pitfalls during graduate school and two postdoctoral labs, but her enthusiasm for the science has kept her going: "You have to like what you are doing most of the time. If you are venting and whining every day, get out!" ■

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

*Some names have been changed to protect privacy.

MOVERS

Anne Glover, chief scientific adviser for Scotland



2001-present: Personal chair, molecular and cell biology, University of Aberdeen, Scotland

1999-2003: Technical director, Remedios, Aberdeen, Scotland

1998-2001: Reader, molecular and cell biology, University of Aberdeen, Scotland

Anne Glover already has plans for when she assumes the role of Scotland's chief scientific adviser next month. She wants science to become part of popular culture. "I'd like citizens to see that everything from probiotic foods to mobile phones is supplied by science," she says.

Her varied background, including a spell in business, should help her in that quest. As a youngster, Glover's interests ranged from chemistry and astronomy through to biology. As an undergraduate, she spent a year studying chemistry at the University of Edinburgh, but switched her degree to the emerging field of biochemistry at the suggestion of her mentor. The move paid off and, after a PhD at the University of Cambridge, Glover found herself doing a postdoc in Aberdeen studying the genome of slime mould.

Her work with slime mould gave her valuable experience in molecular genetics and, after a year spent developing DNA sequencing techniques, Glover was granted a lectureship under the 'new blood' scheme. This UK-wide initiative sought to attract and keep young scientists in academia, largely by reducing their teaching load to allow them to focus on their research.

Glover's move into the world of commerce came about thanks to what she describes as her greatest achievement: genetically engineering microorganisms to make them environmental biosensors. Initially, this work was aimed at tracking the survival of genetically modified organisms released into the environment. "Once you release organisms in to the soil, it's hard to get them back," she says. It was also used to assess whether genes were being exchanged between the modified organisms and the natural community.

To get this information, Glover needed a way of spotting when a modified gene was expressed. She hit upon the idea of modifying it so that it would glow in the dark when active. She later modified the technique so that this 'bioluminescence' occurred in response to changes in the environment, such as the presence of certain toxic compounds.

These biosensors brought out the entrepreneur in Glover and she set up Remedios as a spin-off from the University of Aberdeen to market the technology. In 2000, the industry and government group Biotech Scotland named Remedios as the nation's best new biotech company. Glover has since left the business, but true to her new role and vision, she encourages every scientist to think about the real-world, practical implications of their research.

Virginia Gewin

RECRUITERS & INDUSTRY

From student to entrepreneur

When I joined a chemical-engineering graduate programme after working in the pharmaceutical industry, I wasn't sure whether I would ultimately return to the commercial sector or enter academia. The independence enjoyed by professors was appealing, but I wanted my work to have a direct impact on clinical practice. I've discovered that becoming an entrepreneur has the advantages of both career choices. Entrepreneurs have the freedom to direct technological development and the ability to bring a product to the clinic.

The biotechnology sector intrigued me. I liked the idea of building a small, innovative company. By my third year of research, I had the basis for a new product: an antimicrobial coating for medical devices. I was fortunate to find seasoned advisers with proven track records of translating academic research into commercial applications. This was a crucial step.

To evaluate what resources I'd need to turn my science into a commercial success, I entered the Massachusetts Institute of Technology's \$100K Entrepreneurship Competition last autumn along with my business partner David Lucchino, an MBA student. This spring, the competition's panel awarded our company, SteriCoat, the contest's top

prize. We went on to win Harvard's Graduate School of Arts and Sciences Biotechnology Competition and, in early July, Britain's Oxford Business Plan Competition.

I knew that creating a biotech start-up was inherently risky given that the long development times and clinical hurdles make it difficult to attract investors. I've learned that the keys to success are finding a clear problem with an elegant and compelling solution, and pulling together an experienced team.

Indeed, identifying a clear problem is my top recommendation for young researchers interested in breaking into biotech. Our product, SteriCoat, prevents infections from forming on medical devices such as catheters and heart implants.

Prospective entrepreneurs should be willing to take advice and criticism from potential users and business experts. And they should seek advisers and co-workers whom they trust and enjoy working with. This not only increases the chances of success, but makes the long hours and high-pressure environment much more bearable.

Christopher Loose is a fourth-year chemical-engineering graduate student at the Massachusetts Institute of Technology and future chief technology officer of SteriCoat.

GRADUATE JOURNAL

Lost and found

Moving usually involves making surprising discoveries in cupboards and drawers. While packing last month, I unearthed an ABC book my older sister made for me, several high-school essays, as well as all the notes from my first year at university. Apart from the alphabet, which fortunately I still remember, I was astonished at how much information has vanished from my brain. Apparently, not so long ago, I could compare the philosophers of the Enlightenment, summarize the reasons for the Great Depression, and distinguish between the many types of Finnish bogs — all this is now utterly lost.

But with a second look, my feeling of depression faded. The main messages still seemed familiar. True, I didn't remember all the details, but the essentials from my years of schooling were there, ready to be awakened by a little prompting.

The bundles of paper reminded me about the importance of major ideas and scientific themes. Writing up my thesis means I must delve into the smallest details of social life in an ant colony. But thinking outside the box is often useful and inspiring. To find new approaches to old problems, I like to explore other fields through teaching or attending seminars. Simply associating with students in other disciplines and departments can help. So, for those lacking inspiration, don't just stare at the computer screen. Try a game of beach football with your peers.

Katja Bargum is a graduate student in the Department of Biological and Environmental Sciences at the University of Helsinki, Finland.

From the desk of Jarrod Foster

A brain implant for whatever ails you.

Biren Shah

Sent: 5/28/2009 11:02 AM

From: Jarrod Foster [jfoster@insight-market.com]

To: Powell BioEnhancements

Subject: BIRS Market Study

Attachments: BIRS-Insight.pdf (2.75 MB)

Dear Powell Team:

Insight Market Research's study into the performance of the Brain Implant Reasoning Stimulator (BIRS) 4.0 has shown very promising market penetration in initial segments. BIRS has achieved just under 4% penetration among university faculty and students. The response from private industry has been even better. Nearly 10% of R&D employees in target industries have been fitted with BIRS sponsored by their employers.

Surveyed institutions report significant productivity gains including: shortened times to market, lower development costs and parallel development projects executed by the same workforce. In total, our results show a 25% increase in productivity in BIRS-implemented R&D personnel.

Given this obvious success, we are confident that adoption rates will dramatically increase. We have included several recommendations in our full report to help Powell maximize profits from future sales.

However, it has also come to our attention that even at this early stage in adoption, the security measures implemented have been breached. Hackers have developed illegal software to enable 'off-label' functionality. Although BIRS is designed to increase the reasoning capacities of an individual (as defined in modern theories of bounded rationality), these modifications are generally focused on *reducing* mental capacities.

The most popular hack, called Devo, references 'levels' ranging from 0 to 4, representing the degree of retardation of mental faculties. For most subjects, level 4 is only a minor decrease in mental ability. Level 0 is a full shutdown of higher mental function, effectively an implant-induced coma.

Devo is available for download from many Internet sites. Legal action will not eliminate it, although such action may reduce liability. A risk-management study is needed to determine the degree of Powell's exposure. Powell should consult

with legal counsel to limit liability for any injuries or deaths caused by unapproved uses of BIRS.

A few of the more salient survey responses illustrate some very powerful and unforeseen capabilities of BIRS technology:

"My wife and I both work full time. With the demands of three kids, our jobs, the bills, and our ageing parents, it can be impossible to get a moment together. Even when we do, it's not easy to get over the stress. Our love life has definitely suffered. But BIRS at level 2 lets us forget about the rest of the world when we're alone. The sex



hasn't been this good since we were in college! I definitely think BIRS has saved my marriage. Who knows? It may have saved me from an early heart attack."

— 36-year-old male

"My friends and me went to this frat party and some of them had [the BIRS implants]. They were playing this crazy game... You take someone with it and blindfold 'em and sit 'em down with the music off and everyone really quiet. You've gotta be really quiet for a couple minutes or it doesn't work so good. But then they set themselves to [level] 1 and take off the blindfold and look around. It's like amazing what happens when people are down that low! One guy just started trying to make out with the first girl he saw. She was totally beat and his girlfriend was right there! We had one girl just run off and like hide in the closet yelling like 'Oh my God! Don't touch me!' until some of the guys dragged her out and turned it off. Another guy, he's always such a jerk, but he went around begging for everyone to say they liked him. It's like people get really drunk, but without the sloppiness and puking."

— 20-year-old female

"My brother's been staying with me since his wife died. He's devastated, he loved her so much. They got married last year after a long engagement. They were waiting because he wanted to be able to buy a house for them to move in to. You know, right when they got back from their honeymoon. Carry her over the threshold. In fact, he got [BIRS] because he knew he'd get a raise and afford the house sooner. Now, he's got [BIRS] on a timer. He gets home from work and sets [BIRS] to level 0. The timer shuts it off the next morning so he can go back to work. Until then, it's...it's like he's dead. It's horrible how he lies there." — 28-year-old female

"John Doe found under garbage in an alley near Tropicana. Cause of Death: Suffocation following severe trauma to the throat... Patterns of tissue damage suggest attack by a canine, possibly a dog or coyote... [John Doe] was implanted with [BIRS]... Detectives confirmed that it was set to 'level 0' at the time of death, referencing illegal modifications to the unit... [which may explain] the lack of defensive wounds or evidence of struggle... The victim may have been in the alley for a significant period of time before death. Exactly how long remains under investigation."

— Las Vegas Police Report

Although potentially dangerous, off-label uses have clear appeal to diverse demographics. Devo could take BIRS from the academic and research markets into the mainstream. Depending on the regulatory response, Powell should release a legitimate version of these applications or at least avoid overzealous interference with off-label use. Either way, it is especially time-dependent that Powell apply for patents covering these new applications before they can be classified as 'prior art'.

It has been a pleasure working on this project. I am personally amazed at the potential BIRS has for improving individual performance. I have submitted a proposal asking management to invest in implants for all research staff here at Insight.

Sincerely,

Jarrod Foster, Project Lead

Biren Shah is an entrepreneur and aspiring writer. He can be found online at www.birensah.com and in person in Los Angeles, California.

JACEY